TECHNOLOGICAL UNDERSTANDING AND ETHICS READINESS OF TECHNOLOGY



ELINE DE JONG

TECHNOLOGICAL UNDERSTANDING AND ETHICS READINESS OF TECHNOLOGY

ELINE DE JONG

ILLC Dissertation Series DS-2026-12



INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Science Park 107 1098 XG Amsterdam
phone: +31-20-525 6051
email: illc@uva.nl
homepage: http://www.illc.uva.nl/

This publication is part of the project Quantum Impact on Societal Security (NWA.1436.20.002), funded by the Dutch National Science Agenda (NWA) of the Dutch Research Council (NWO)

Copyright © 2026 by Eline de Jong
Cover picture by Anna Tarazevich
Cover design by Eline de Jong
Printed and bound by ProefschriftMaken
ISBN: 978-94-6534-667-0

Technological Understanding and Ethics Readiness of Technology

The role of understanding and timing in the ethics of quantum technologies

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op donderdag 3 december 2026, te 16.00 uur

door Eline Louise de Jong
geboren te Amsterdam

Promotiecommissie

Promotores:	prof. dr. S.J.L. Smets	Universiteit van Amsterdam
	dr. S. De Haro Ollé	Universiteit van Amsterdam
Copromotores:	mr. dr. B.F.W. Wernaart	Fontys Hogescholen
	dr. C. Shelley-Egan	Technische Universiteit Delft
Overige leden:	prof. dr. ing. R.A.M. van Rooij	Universiteit van Amsterdam
	prof. dr. H.O. Dijstelbloem	Universiteit van Amsterdam
	prof. dr. C. Schaffner	Universiteit van Amsterdam
	prof. dr. J.V.J. van Hoboken	Universiteit van Amsterdam
	prof. dr. H.W. de Regt	Radboud Universiteit
	prof. dr. A. Grunwald	Karlsruhe Institute of Technology
	dr. J. Cramer	Universiteit Leiden

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

To Nellie and Bonnie,

for all the wisdom and joy you bring into my life

Table of contents

Acknowledgements ..... **i**

Preface ..... **v**

Chapter 1. Introduction **1**

1. Research question 2
2. A brief introduction to quantum technologies 3
   - 2.1 What are quantum technologies? 3
   - 2.2 The emerging ethics discussion of quantum technologies 6
3. The ethics of new and emerging technologies: an epistemic perspective 10
   - 3.1 What are new and emerging technologies? 10
   - 3.2 Understanding: coming to grips with novel technology 11
   - 3.3 Timing: engaging in ethics upstream 12
4. Chapter overview 14

Part I. Towards a Theory of Technological Understanding *21*

Chapter 2. A Conception of Technological Understanding **23**

Abstract 23

1. Introduction 24
2. Setting the stage 27
   - 2.1 Scientific understanding 27
   - 2.2 A definition of technological artefacts 29
   - 2.3 An analogy between scientific theories and technological artefacts 34
3. A conception of technological understanding 36
   - 3.1 Technological understanding: the ability to use a technological artefact 37
   - 3.2 Specifying technological understanding: three logical options 41
   - 3.3 Technological understanding in a context of design 43
4. Conclusion 46

Chapter 3. A Contextual Approach to Technological Understanding **49**

Abstract 49

1. Introduction 50
2. The context-dependence of technological understanding 51

2.1 Technological understanding 51
2.2 The pragmatic character of technological understanding 54
2.3 Three contexts of technological understanding 55
3. Specifying technological understanding: three types 60
3.1. Design-type technological understanding 60
3.2. Operation-type technological understanding 61
3.3 Innovation-type technological understanding 63
4. Assessing technological understanding 67
4.1 Technological understanding as a form of inferential reasoning 67
4.2 Inferential abilities per context 70
5. Conclusion 73

Chapter 4. Technological Understanding and Ethics 77

Abstract 77
1. Introduction 78
2. Understanding as a prerequisite for ethical discussion 79
2.1 More than knowledge 79
2.2 Three types of technological understanding 80
3. A functional understanding of technology 82
3.1 From physics to functions 82
3.2 Implications for ethical engagement with (quantum) technologies 85
4. Conclusion 86

Part II. Introducing the Ethics Readiness of Technology 89

Chapter 5. Ethics Readiness of Technology 91

Abstract 91
1. Introduction 92
2. Outcomes- and meaning-oriented approaches to ethics of technology 93
2.1 Upstream ethics and the problem of uncertainty 93
2.2 Outcomes-oriented ethics: engaging with possible futures 95
2.3 Meaning-oriented ethics: assigning meaning to future technologies 97
3. Evaluating ethical approaches: adding a temporal dimension 100
4. The ethics readiness of technology and its implications for ethics 101
4.1 Technology readiness and societal readiness 102
4.2 Ethics readiness 104
4.3 Pairing technology readiness with ethics readiness 105
5. The ethics readiness of quantum technologies 108
6. Conclusion 110

Chapter 6. Visions of Quantum Computing 113

Abstract 113

1. Introduction 114

2. Method: Ethical Vision Assessment 116

2.1 Stage one: identifying visions 116

2.2 Stage two: ethical analysis 122

3. Visions of quantum computing 123

3.1 Findings: value-oriented language, domains, applications, and technical affordances 123

3.2 Quantum computing and the promise of change 130

3.3 Four visions of quantum computing 132

4. Ethical analysis of visions of quantum computing 135

4.1 Content (substantive level): techno-optimism and technosolutionism 136

4.2 Framing (discursive level): the good lies in the future and requires innovation 139

4.3 Authors (agential level): interest shaped narratives 140

5. Conclusion 142

Chapter 7. Technosolutionism and the Case of Quantum for Good 145

Abstract 145

1. Introduction 146

2. The emergence of Quantum for Good 147

2.1 Quantum for Good as a normative ambition 148

2.2 In search of 'use cases' 149

3. Technosolutionism: from pejorative term to critical lens 150

3.1 Critical accounts of technosolutionism 150

3.2 Nuanced views of technosolutionism 152

4. Ethical assessment of technosolutions 155

4.1 Guiding questions for ethical scrutiny 156

5. Assessing Quantum for Good through a lens of technosolutionism 157

5.1 Problem construction 163

5.2 Sociopolitical dimensions 165

5.3 Scope 166

5.4 Legitimacy 168

6. Conclusion 169

Chapter 8. Risk and Precaution in the Quantum Discourse 173

Abstract 173

1. Introduction 174

2. The quantum cyber security threat, its solutions and its challenges 177
2.1 Breaking widely-deployed encryption 177
2.2 Transitioning to quantum-safe cryptography 179
3. Risk and precaution as part of technological visions 181
3.1 A philosophical perspective on risk 181
3.2 The principle of precaution 182
3.3 Ethical-social dimension of risk and precaution: a vision approach 185
4. A framework for assessing risk and precaution as a vision 188
4.1 Three levels of analysis 188
4.2 Evaluating questions 189
5. Migrating to quantum-safe cryptography: Better secure than sorry? 191
5.1 The risk: a cryptographically-relevant quantum computer 191
5.2 The precautionary response: implementing quantum-safe cryptography 196
5.3 Plausible risk, appropriate precaution? 201
6. Conclusion 202

Chapter 9. Conclusion and discussion 207

1. The epistemic conditionality of ethics 207
2. Research questions 207
3. Sufficient understanding of technology 209
4. Appropriate timing of ethics 211
5. Reflections and further research 214
5.1 Division of epistemic labour 214
5.2 Division of moral labour 215
6. Closing remarks 218

References 219

Technological Understanding and Ethics Readiness of Technology 245

Technologisch Begrip en Ethische Ontvankelijkheid van Technologie 251

Abstract 259

Appendix 261

Other publications by the author 263

Titles in the ILLC Dissertation Series 267

Acknowledgements

If a word of acknowledgement is meant to reflect on what and who enabled me to deliver this thesis, I should start by acknowledging my privileged position, living in a place where I take freedom for granted, where I can freely express myself, and where I get paid for developing my thoughts. I realise that, at the very same time, many humans and non-human animals live in fear, oppression, and pain, and will never get the chance at life that I got by sheer luck. I owe it to them, not only to acknowledge my privilege, but to ensure that in the next steps of my career I use this position to pursue justice for those who are wronged.

The University of Amsterdam's ILLC, IoP, and QuSoft gave me the opportunity to conduct this research and the freedom to follow my curiosity and personal engagement. I thank my supervisors, Sebastian De Haro, Clare Shelley-Egan, and Bart Wernaart, and my promotor, Professor Sonja Smets, for their support throughout this journey. I especially want to thank you, Sebastian, for your engagement with my research, which made me one of the rare PhD candidates who could always count on timely and careful feedback, for sharing your experience with me, and for the trust and space you gave me to give talks across the world, explore new collaborations, and pursue topics beyond quantum. I'm equally grateful to you, Clare, for your support in my personal development and for the chats that always gave me fresh energy, for slowing me down when I was inclined to move too quickly, and for inviting me to hop on your thought trains and publish together.

Then there was Ailsa Robertson, my closest colleague who became a dear friend. I have called you a woman of surprises because you have so many passions and talents – but I think your biggest talent is your loyalty and attentiveness as a friend. Thank you for the many times you lent me a listening ear, for sharing your enlightening view on faith, for all the little postcards you wrote to let me know that you were thinking of me, for enriching my vocabulary with Scottish idioms, and for showing me that adding a little British gentleness to my communication can do miracles.

Over the years, I have met many people working on the ethical, legal, and social aspects of quantum technologies. I have the feeling that we became a real community, learning from each other and together working towards a responsible quantum ecosystem. In particular, I owe gratitude to Mauritz Kop, who was the one who invited me into the quantum realm and believed in me from the start. Thank you, Mauritz, for the opportunity to co-author several articles with an impressive international team, among whom I would especially like to mention Professor Urs Gasser and Professor Timo Minssen. My visits to Stanford were without a doubt among the highlights of my time as a PhD researcher. I would also like to thank Professor Christian Schaffner, Professor

Pieter Vermaas, Julia Cramer, Professor Joris van Hoboken, Zeki Seskir, Adrian Schmidt, Thijs Latten and María Palacios Barea – key figures in the quantum community – for all the discussions we had, which shaped my thinking.

I would also like to thank my previous colleagues of the Netherlands Scientific Council for Government Policy (WRR), where I learned a lot about sound scientific research and the value of peer feedback. Professor Frans Brom and Professor Haroon Sheikh, in particular, have played an important role for me. Frans, thank you for opening the doors of the WRR for me (twice), for your sometimes painfully honest feedback, always balanced by crediting good work, and for coaching me through my first steps towards animal ethics. And Haroon, you have been one of the most brilliant people I have worked with. I have learned so much from you, as the erudite philosopher and strategic mastermind that you are. Thank you for the responsibilities you entrusted me with, and for the friendship we developed.

A special word for Julia, who has taken on so many roles in my life and became my shadow supervisor. I am deeply grateful for your relentless support in everything I do, for the humour you bring into a world that sometimes takes itself too seriously, and for our discussions about the big questions in life. You have been, and continue to be, a true inspiration to me.

I am also grateful to Lucia and Thomas, my parents-in-law, who are so much more than parents-in-law. You not only ask what I've been up to lately, but you also actually listen when I try to make sense of what I'm doing, and you always assure me that I am right and the world is mad (I see that as an act of love). Thank you for taking me into your family without reservation.

A big shout out to my family and friends, who have been there celebrating the highs with me and helping me get through the lows. A special spotlight for Pauline, Hanna, Rikke, Sarah, Saskia, Anna, Max, Anne, Claire, Jonas, Maxime, Mirthe, Pol, Charlotte, Floor, Kyra, Alessandra, Nienke, Floris: thank you for listening to my rants, for reminding me of who I am outside of work, for the long dinner conversations about all things great and small, and for being part of the moments where I forget the time. Thank you for calling, inviting, and checking in on me, even though (or precisely because) I'm bad at doing that myself. I feel carried by you.

I owe everything to my parents, Claudia and Peter. I owe my life to you, obviously, but also the opportunities that got me to where I am today. Dear pap and mam, I know you are proud of me, and that you always play down the role you played in what you are proud of. I think you underestimate what it has meant for me to grow up in a family that always believed in me without projecting expectations onto me. You gave me all the freedom to explore my own path and remained the home I could always retreat to. Your work ethic and your habit of asking critical questions have shaped me as a researcher and underpin this book. But you also passed on a form of sensitivity

that shows in your kindness to animals, in treating them as the personalities that they are. You taught me how to care, and that even the smallest creatures deserve it. Now that I have messed around long enough in the pretentious world of human affairs, I am committed to following my heart, and I know you will support me in my next steps. Parenting is a never-ending project, I'm afraid. But I am extremely grateful to have the most loving, funny, and down-to-earth parents I could have wished for. I'm proud of *you.*

Finally, my gratitude to Olivier is endless. You are my partner in everything, and the person whose view I trust most in the world. Thank you for all the times you checked my PowerPoint presentations, for letting me pick your brain on new ideas for an article when we were supposed to be having a romantic dinner, and for asking every evening how my day was. I'm grateful for your constant support and for your encouragement to pursue my passions, even if that means extending our family with two dogs or trying to earn money with philosophy. I can never thank you enough for always being there for me. I know that this can be a thankless role, sometimes. Heroic, certainly, but not always fun in practice. That makes me appreciate you even more, and I am deeply grateful to have you by my side. I cannot wait to start a new chapter, if only because it is with you.

Preface

In early 2022, I was asked to write an article about this new kid on the block: quantum technologies. More specifically, the suggestion was to apply a framework I had developed together with colleagues at the Scientific Council for Government Policy (WRR) for artificial intelligence (AI) to quantum technologies. "But I have no understanding of quantum technologies whatsoever," I replied. "Don't worry," I was told, "nobody does."

It was during my research on AI that I first noticed how the degree of understanding one has of a technology shapes the questions one asks about its implications. Having little understanding of what AI is, and what it can and cannot do, paves the way for speculations so wild that we might question whether they deserve our attention – at least not yet. Especially because that attention is limited, and because there are more immediate issues that warrant it. By looking so far ahead, we risk overlooking what is urgent now.

I became fascinated by how we can ask meaningful questions about technologies we are only starting to understand. Determined to find my way into quantum technologies and ground my engagement in reality rather than fantasy, I began acquainting myself with this new technological family. In contrast to AI, little material was available to make 'quantum' accessible to non-experts. I struggled through physics papers, and when I asked physicists 'in the wild' to explain what quantum technologies actually are, I quickly found myself lost in translation.

I realised that a physicist's way of understanding would never be mine, not even in simplified form. Still, some understanding of what quantum technologies are and may do seemed necessary to identify the ethical questions they raise. And if understanding is a condition for ethical engagement, it struck me that the same must hold for other ways of engaging with technology – designing it, using it, regulating it. Each of these, in its own way, requires making sense of what a technology is and can do. This became the starting point of my research. I took a few steps back to ask: what does it mean to understand a technology well enough to engage with it meaningfully?

In investigating this question, I stumbled into a further challenge. If a certain kind of understanding is required for meaningful ethical engagement, what does this mean for technologies not yet developed enough to yield such understanding? There is currently little concrete substance about what quantum technologies are capable of for ethics to take a solid bite into. I had argued that understanding is what keeps ethical engagement grounded – but what happens when the understanding we need is limited

by the uncertainties surrounding a new technology? This question further shaped my research.

Over the past years, I have watched a discourse emerge around the ethical dimensions of quantum technologies. This discourse is driven by the ambition – sharpened, perhaps, by our recent experiences with AI, where many ethical issues surfaced only once the technology was already out there – to 'get it right the first time'[1]. I sense eagerness in the ethical discussion of quantum technologies. Yet it is also struggling to genuinely take off, circling the same general concerns without deepening them. This is not a shortcoming of the debate, in my view, but a consequence of its moment. The two concepts developed in this thesis – technological understanding and ethics readiness – are intended to illuminate and address this struggle, and to help identify which questions are worth asking now.

[1] This ambition echoes an ethos from the ethics of nanotechnology, where 'getting it right from the start' became a rallying call (see, for example, Balbus et al., 2002).

Chapter 1

Introduction

> "It is not the case that we have failed to care for Creation, but that we have failed to care for our own creations. We blame the monster, not the creator, and ascribe our sins against Nature to our technologies. But our iniquity is not that we created our technologies, but that we have failed to love and care for them. It is as if we decided that we were unable to follow through with the education of our children."
>
> Bruno Latour, in *Love your monsters* (2011, p. 20)

First we shape our tools and then our tools shape us.[1] You have probably heard this before. Probably too often. Yet I cannot think of a quote that sets the scene for a philosophy and ethics of technology as concisely. In all its simplicity, it captures two things at once.

First, *we* build our technologies. Technological development is not some autonomous force that overcomes us. We invent it, we are its creators. And creation, as Latour (2011, p. 20) reminds us, carries responsibility: bringing something into being makes us answerable for what it becomes.

Second, our technologies shape us in return. With every new technology, the world around us changes. New options, new risks, new ways of living and thinking. Technologies are not mere tools; they mediate our being and reshape the societies that adopt them (Verbeek, 2005).

The first aspect opens up a space for ethics: because we are the ones who develop technologies, we have the power to shape them. The second aspect highlights the ethical stakes of that power: if technology shapes the world we live in, and if we are the ones who shape technology, then we bear a responsibility to consider carefully how we do so.

The recognition that technologies reshape the world and confront us with new realities requiring ethical reflection has given rise to ethics of technology as a subfield of philosophy (Verbeek, 2010). If technology affects the fabric of society, we face the question of how to weave it in well.

[1] The quote is often misattributed to media theorist Marshall McLuhan, but it was actually written by John Culkin, a Professor of Communication at Fordham University and friend of McLuhan, appearing in a 1967 article about McLuhan in the Saturday Review (Culkin, 1967, March 18: pp. 51-53, 71-72). It is, however, fully consistent with McLuhan's thinking. The idea is McLuhan's, the phrasing is Culkin's.

In this thesis, I investigate this question in relation to one particular group of technologies: quantum technologies. Because of their potential impact on research and society, quantum technologies invite ethical engagement. Moreover, being at an early stage of development, they exemplify the opportunity *and* challenge of engaging with the ethical aspects of a new technology before it has fully taken shape. Quantum technologies therefore serve not only as a direct subject of ethical enquiry but also as a paradigm case for the challenges of any ethics of new and emerging technology. But before asking what is at stake as quantum technologies move from lab into society, this thesis examines what is needed to ask that question meaningfully. What enables meaningful ethical engagement with technology, especially new and emerging technologies?

I take an ethical question to be meaningful if it is grounded in technological reality, getting both its substance and urgency from what the technology is capable of and what can plausibly be expected of it. Without such grounding, ethical discussion risks raising misguided or premature questions, misdirecting valuable attention. Meaningful ethical engagement is instead genuinely informed by what the technology is and may do, to the best of our understanding. It engages with issues that are relevant and tractable at a given stage of development. In other words, it is about asking the *right questions* at the *right moment*.

1. Research question

Articulating what enables meaningful ethical discussions about technology seems important for fostering such discussions, yet it remains largely absent from the ethical discourse on technology. A vast number of studies focus on applied ethics, many reflect on the challenges of technology ethics (e.g. Nordmann, 2007; Sollie, 2009; Manders-Huits, 2011; Brey, 2012), others propose methods to address these challenges (e.g. Reijers et al., 2018; Hansson, 2017), and some hint at the requirements of grounded ethical analysis (e.g. Nordmann & Rip, 2009; Lucivero et al., 2011). The conditions for meaningful ethical discussion of technology, however, have not yet been systematically worked out.

The question about enabling conditions for meaningful ethical discussion is especially relevant in the context of new and emerging technologies such as quantum technologies. As a family of largely early-stage technologies, engaging with the ethical aspects of quantum technologies comes with two related challenges: their novelty demands that we get to grips with what they are and what they may do, while their early-stage development leaves much uncertain – about when they will materialise, how, and with what implications.

These aspects illuminate the epistemic dimension of ethics: what we know – and what we do not – has direct bearing on the ethical questions we can meaningfully raise.

There is thus a direct connection between the epistemic question of what we can know and the ethical question of what we ought to do. The literature on the ethics of technology gestures towards this epistemic dimension in at least two ways: in its recognition of uncertainty as a central challenge for early-stage ethics (Sollie, 2007), and in its calls for improving understanding of specific technologies to foster ethical engagement (Vermaas, 2017; WRR, 2021). However, no systematic account of what is required, epistemically, for a meaningful ethical discussion of technology has yet been developed. This motivates the core question of this thesis:

> *What are the epistemic conditions for meaningful ethical engagement with technology?*

This thesis argues that the ethics of new and emerging technology such as quantum is characterised by a twofold epistemic challenge. It requires coming to grips with what a technology is and may do, and navigating the deep uncertainties surrounding its development. I investigate what those requirements entail and how they enable meaningful ethical discussion: What does it mean to sufficiently understand a technology for ethical engagement, and how can we avoid problematic speculation without being pushed into a purely reactive position? These questions structure this thesis.

Beyond developing the concepts and frameworks that help to navigate these challenges, this thesis also puts them to work in engaging with ethical questions raised by quantum technologies.

2. A brief introduction to quantum technologies

2.1 What are quantum technologies?

Harnessing the principles of quantum mechanics, a new family of technologies is taking shape: quantum technologies. In the plural, because there are multiple sorts. What they share is that they use quantum mechanics to achieve capacities beyond those of technologies based on classical physics. To refresh the memory: quantum mechanics is a highly successful theory in physics that describes how nature – as matter and energy – behaves at the smallest scales, that is, at the level of atoms and subatomic particles. At this level, phenomena occur that we do not recognise in our everyday experience and that are sometimes counterintuitive to the rules of physics we are naturally familiar with. For example, at a quantum level a particle can exist in a combination of distinct states – a coin being heads *and* tails or a bit being 0 *and* 1 – simultaneously (superposition) (Hoofnagle & Garfinkel, 2022: pp. 27-9) – something we can barely picture. Or two particles can become linked such that their properties remain

correlated even when they are far apart (entanglement) (pp. 26-7) – something we would rather associate with spirituality or science fiction than with serious physics.

Such phenomena are leveraged in the field of quantum technology to develop new classes of technology. A first generation includes Magnetic Resonance Imaging (MRI), lasers, and transistors. These technologies take advantage of natural quantum mechanical processes without controlling specific quantum behaviour. More recently, a second generation has gained momentum, aimed at the *controlled* utilisation of quantum phenomena. It is this deliberate manipulation that underpins new classes of quantum technology:

Quantum sensing harnesses quantum mechanics to measure physical information with exceptional precision. The extreme sensitivity of quantum sensors can achieve measurement accuracy far beyond what classical sensors allow. Applications include highly accurate navigation systems that do not rely on GPS – useful for submarines and in GPS-denied environments (Pei et al., 2025), early detection of gravitational anomalies for geological surveying or mineral exploration (Lin et al., 2022), and medical imaging that can detect brain activity or tumours with greater sensitivity than current technology (Aslam et al., 2023). Of the second-generation quantum technologies, quantum sensing is the most mature: several applications are already operational or in advanced stages of development.

Quantum computing uses quantum mechanical phenomena to process information and perform computational operations. Unlike classical computers, which rely on the physics of our everyday world, quantum computers harness quantum phenomena to achieve distinct computational capabilities. This does not mean they will replace your laptop or simply put existing computers 'on steroids'. Rather, they promise to make certain problems tractable that remain beyond the reach of even the most powerful classical supercomputers. A near-term and particularly promising application is quantum *simulation*: using quantum systems to model complex physical and chemical phenomena that are intractable for classical computers, with potential contributions to materials science (Paudel et al., 2022) and drug discovery (Cao et al., 2018). More broadly, quantum computing promises exponential speedup for certain classes of problems, with applications including the optimisation of complex logistical systems such as supply chains or financial portfolios (Weinberg et al., 2023; Phillipson, 2024), and breaking – or stress-testing – current encryption methods (Buchanan & Woodward, 2017). Textbox 1 discusses the latter in more detail. Quantum simulation is more limited in scope and therefore more near-term; general-purpose quantum computing at scale remains considerably more speculative.[2]

[2] Quantum simulation and computing are sometimes treated as distinct classes, with simulation referring specifically to using a quantum system to model a natural quantum system for scientific purposes, and

Quantum communication leverages quantum phenomena to transmit information over distance. Its promise lies not in faster communication but in fundamentally more secure communication, guaranteed by the laws of physics rather than by computational assumptions. The currently most investigated application is quantum key distribution (QKD), which allows two parties to share an encryption key such that any interception is detectable (Gisin & Thew, 2007). Longer term, quantum communication is expected to form the backbone of a quantum *internet*: a network enabling secure communication and distributed quantum computing across distances (Luo et al., 2023). Quantum communication sits between sensing and computing in terms of maturity: quantum key distribution systems exist and are being tested in real-world settings (Tippmann et al., 2026), but widespread deployment and a real quantum internet remain longer-term prospects.

Textbox 1. Quantum computing and its threat to cyber security

A cryptocalypse?

Among second-generation quantum technologies, quantum computing probably attracts the most attention. Emerging in the 1980s at the intersection of quantum mechanics and computer science, the field has progressed from theoretical foundations to early experimental implementations, with considerable advances in recent years. Instead of working with the binary digits ('bits') of classical computers, quantum computers work with quantum bits, or 'qubits'. Harnessing quantum phenomena, this promises to make certain computations far more efficient than classical computers can achieve. In particular, quantum computers offer the prospect of an exponential speed-up for problems whose complexity scales exponentially with size, making previously infeasible computations potentially achievable within practical timeframes (Preskill, 2018).

Although scalable quantum computers remain elusive, the field is widely regarded as holding enormous potential, as reflected by the significant private and public investment in its research and development. Since the first experimental steps towards building quantum computers in the late 1990s and early 2000s, major technology firms – including Google, IBM, and Microsoft – are now investing heavily in this field. Estimates place the potential economic value of quantum computing between $900 billion, and $2,000 billion by 2035 (McKinsey, April 2024). Government interest has also surged, with national strategies fuelling a global race to develop practical quantum computers. More recently, alongside commercially and strategically driven efforts, socially motivated initiatives have emerged, seeking to harness quantum computing to address grand societal challenges (Quantum for Good, n.d.; Open Quantum Institute, n.d.).

computing referring to more general-purpose information processing. In practice, the boundary is blurry, as simulation can be understood as a more limited application of quantum computing.

Where exactly the potential of quantum computing lies remains an open question. The expectations range from limited but significant applicability in simulating natural quantum systems, contributing to fundamental physics and chemistry, to groundbreaking applications in healthcare, materials science, finance, defence, logistics, clean energy, machine learning, and security. A remark by physicist John Preskill captures the open-endedness well: "My gut feeling is that if powerful quantum computers were available, we would somehow think of many clever ways to use them." (Preskill, 1998). Put differently: the applications will reveal themselves once the technology arrives.

One important application of quantum computing – and one that explains much of the attention it receives – is breaking encryption. Encryption is the locking of information so that it is accessible only to authorised parties. Or put differently: it is a way of protecting data from unauthorised access. It works by relying on mathematical problems so complex that even the most powerful supercomputers cannot solve them within any practical timeframe; the computational difficulty is, in effect, the lock. However, recall that one of the promises of quantum computing is solving precisely such problems. Now, if large-scale quantum computers can do this, a vast range of current encryption methods[3] will become obsolete overnight, leaving private and sensitive information – from personal messages and bank transfers to business and state secrets – freely accessible. The implications for our digitalised world hardly need spelling out. Q-day – the projected moment a quantum computer capable of breaking current cryptography becomes operational – is not without reason by some dubbed the "cryptocalypse" (Lindsay, 2020). Chapter 8 discusses this scenario in more detail.

2.2 The emerging ethics discussion of quantum technologies

Advancing research and development, and a growing sense of what quantum technologies might be capable of, prompted a discussion about the implications of their step from lab into society. Vermaas' special issue on the societal impact of quantum technologies (2017) gave an early impetus to this discussion (Vermaas, 2017; De Wolf, 2017; Möller & Vuik, 2017). Several studies emerged examining the ethical, legal, social, and policy aspects of quantum technologies (World Economic Forum, 2022; de Jong, 2022; Kop, 2023; Possati, 2023; Umbrello, 2024; Shelley-Egan, 2025; Shelley-Egan & de Jong, 2025; Nguyen, 2025; Barea, 2026).

[3] The quantum cyber security threat primarily concerns asymmetric encryption schemes such as RSA and ECC, whose security depends on mathematical problems that a large-scale quantum computer could solve efficiently. However, not all cryptographic methods are equally affected. Symmetric encryption, for example, relies on a different type of mathematical structure and is therefore comparatively resilient: although quantum computing would weaken it by making certain attacks faster, the loss of security can generally be addressed relatively easily (see also Chapter 8, footnote 7).

Alongside these discussions, a more normative question emerged: how to promote the 'right' impact (Von Schomberg, 2013, 2014). *Responsible innovation* became a central concept in this debate, with many studies examining what responsible innovation in the quantum realm entails and how it should be pursued (Von Schomberg, 2013; Owen et al., 2013; Inglesant et al., 2021; Coenen & Grunwald, 2017; Ten Holter et al., 2023; Perrier, 2021; Kop et al., 2024a, 2024b; Gasser, de Jong, & Kop, 2024; Van der Enden et al., 2024).

More specifically focusing on the ethical aspects of quantum technologies, an ambivalent picture emerges (Shelley-Egan & de Jong, 2025). Quantum technologies may have positive effects on public values – values with broad societal relevance – as well as negative ones. Many applications have a dual-use character in both a narrow and a broad sense: in the narrow sense, they may serve civilian as well as military ends; in the broad sense, the same capability can be both beneficial and harmful depending on the purposes for which it is deployed. This ambivalent character cuts across a number of ethical domains, of which the most prominent are security, privacy, and sustainability. I start with sustainability; often discussed last, yet capturing a significant and frequently underappreciated ethical dimension of quantum technologies.

Sustainability – On the one hand, the capabilities of quantum technologies may contribute to sustainability goals, with respect to the environment and beyond. They are increasingly mentioned alongside the United Nations Sustainable Development Goals, with proponents stressing their potential to contribute to addressing climate change, improving energy efficiency, and enabling better climate modelling and monitoring (Open Quantum Institute, n.d.; World Economic Forum, 2024a; Root, 2025). More concretely, quantum computing may contribute to the design of better batteries, carbon capture technologies, and the optimisation of energy grids. On the other hand, there are concerns about the sustainability of quantum technologies themselves. While they may reduce energy consumption for certain computational tasks due to higher efficiency, they also require energy in new ways, for example for cooling their systems (Jaschke & Montangero, 2023; Meier & Yamasaki, 2025). Moreover, if quantum technologies will be deployed in superimposition on existing technologies and infrastructure rather than replacing them, their energy demands will add to, rather than offset, the considerable energy consumption of the current digital ecosystem. Meanwhile, the production of quantum hardware may depend on rare earth materials that are already under pressure, and whose extraction carries significant environmental costs (World Economic Forum, 2024b).

Privacy – When it comes to privacy, quantum technologies present both opportunities and risks. On the opportunity side, quantum communication may enhance the protection of personal data through quantum-enhanced encryption methods, known as quantum key distribution (QKD), and quantum networks that make unauthorised

interception immediately detectable. On the risk side, quantum computing introduces a new privacy threat by its ability to break encryption methods currently used to protect personal data. Enhanced sensing capabilities enable the collection of more and more precise data, while quantum computing's expanded capacity to process large datasets could amplify existing surveillance capabilities. This raises questions about new and amplified threats to data privacy and citizen surveillance (De Wolf, 2017; Malina et al., 2021; World Economic Forum, 2024b; Abdikhakimov, 2024).

Security – Regarding cyber security, quantum computing introduces a significant risk by its potential to break widely used encryption methods, with implications ranging from the exposure of personal data to the compromise of sensitive business information and state intelligence (see box 1). This threat may already apply now in the form of harvest-now-decrypt-later strategy (Singh, 2024): adversaries may already start collecting encrypted information, anticipating that they will be able to access it when a large-scale quantum computer becomes available. At the same time, quantum-based techniques such as QKD may improve security by making unnoticed interception of communication physically impossible. Beyond cyberspace, quantum technologies present a similarly dual picture for broader security. In the military domain, enhanced sensing and computing capabilities cut both ways: they improve the precision and effectiveness of military technology, serving as both threat and opportunity depending on whose hands they are in (Krelina, 2021; Nahar, 2023). Defence considerations remain among the most prominent drivers of government investment in quantum technologies (Vogiatzoglou, 2025; see also Inglesant et al., 2018; Sayler, 2026).

Overarching these domain-specific concerns are questions of *justice and fairness*, operating at two distinct levels. At the distributional level, questions arise about fair access to the capabilities of quantum technologies and the equitable distribution of their benefits and harms: who benefits, and who bears the risks? This question, as Barea (2026) rightly foregrounds, cannot be seen apart from questions about justice at the procedural level: who shapes these technologies in the first place? The development of quantum technologies is currently driven primarily by a small group of powerful actors – (large) technology companies and a handful of (wealthy) states – and a research and development community that remains predominantly male and from the global north (Sainz, 2022; Beige et al., 2025; Propp et al., 2026) and is generally highly educated. This raises questions about whose stakes, worldviews, and blind spots shape the innovation trajectory, and by extension, which values become embedded in the technology and whose interests it will ultimately serve. Such concerns underpin growing warnings of a 'quantum divide' emerging at both local and global levels (Ten Holter et al., 2022; World Economic Forum, 2024b; Gercek & Seskir, 2025; Schmidt et al., 2025). As Barea (2026, p. 9) warns: "If unchecked, [quantum technologies] may yield a 'quantum divide', where benefits accrue to a privileged minority while global inequities deepen."

Reflecting on this emerging ethical discourse, two observations can be made. The first is that many of these issues are not specific to quantum technologies. We know them from other technologies, and they now present themselves in a new context. This does not, however, make them any less relevant. The fact that an issue is not novel does not mean it has been addressed. We do not need a specific 'quantum ethics' to have a meaningful discussion of the ethical issues arising in the context of quantum technologies. On the contrary, Sætra and Danaher (2022) warn against the fragmentation of technology ethics into ever-narrower domain-specific silos, driven by a "zeal for carving out a new domain-specific ethics" (p. 93). Instead, they argue that there is genuine value in recognising domain-specific ethical discussions as instances of general issues in technology ethics. Doing so allows us to draw on the accumulated insights of broader technology ethics, rather than reinventing the wheel each time a new technology emerges.

In an article that is not part of this thesis, Clare Shelley-Egan and I make a similar argument in the context of quantum technologies, warning against the 'allure of novelty': the risk of focusing mainly on what is unprecedented while overlooking continuities with earlier technological developments and the ethical discussions they raised (Shelley-Egan & de Jong, 2025). The fact that an ethical concern is familiar does not mean it is settled. Moreover, in a new technological context, a known issue may intensify, take new forms, or interact with the technology's specific features in ways that require renewed scrutiny. It is thus not only legitimate but also ethically imperative to attend to issues that may not be quantum-specific but yet quantum-relevant. This includes questions about equitable access to the benefits of quantum technologies, hype and inflated promises, technosolutionism, power concentration, dual use, and responsible innovation. In addition to remaining alert to genuinely novel ethical issues, we therefore call for being attentive to how familiar ethical concerns resurface in new contexts.

The second observation is that many of the applications on which these ethical issues centre are still speculative. This applies primarily to quantum computing and communication, which remain in early experimental stages. In the absence of full-fledged technologies – and of evidence that clearly establishes their feasibility – the ethical issues raised are attributed to what Lucivero et al. (2011, p. 129) call 'elusive objects': technologies whose future form and capabilities remain fundamentally uncertain. Such uncertainty renders the applications and impacts of these technologies a fertile ground for speculations. The field of quantum is not shielded from hype dynamics either: speculative promises and alarmist scenarios set the agenda, both resting on contestable assumptions (Smith III, 2020; Ezratty, 2022; de Jong, 2022, p. 15; Barea, 2026, p. 6). If anything, this should make ethics epistemically humble and cautious (Nordmann & Rip, 2009): careful not to let speculative promises set the agenda, and attentive to the assumptions on which ethical concerns rest.

Shelley-Egan and I (2025) flag the uncertainties around quantum technologies as a continuity with the ethical discussion of nanotechnology in the late 2000s. The latter was similarly characterised by a deep uncertainty about the materialisation of the technology. Swierstra and Rip (2007) highlight this speculative character as a recurring aspect – and potential pitfall – of what they call 'NEST-ethics': the ethics of new and emerging science and technology. They note the related difficulty of acting under conditions of ignorance as the central challenge of this field. In doing so, they – and others recognising this challenge (e.g. Nordmann, 2007; Sollie, 2007; Lucivero et al., 2011; Grunwald, 2011; Brey, 2012) – bring into view the epistemic dimension of any ethics of NEST: engaging with these technologies means operating under a specific epistemic condition, characterised by a *new* object surrounded by *uncertainty*. Yet a systematic account of this epistemic dimension, one that clarifies what the relationship between knowledge and ethics entails beyond recognising uncertainty as a challenge to be managed, remains lacking.

This thesis seeks to address this gap by investigating the epistemic conditions for meaningful ethical discussion of technology and putting the resulting account to work in the emerging ethical debate about quantum technologies.

3. The ethics of new and emerging technologies: an epistemic perspective

3.1 What are new and emerging technologies?

Just as technology ethics is a subfield of philosophy of technology, the ethics of new and emerging technologies is a specific focus within technology ethics. It is distinguished by its concern with early-stage technology and the opportunity for early-stage intervention (Brey, 2017). Brey defines emerging technology as follows:

> "Emerging technologies are technologies that are new, innovative, and still in development, and are expected to have a large socioeconomic impact. They are new in the sense that they employ new concepts, methods and techniques and cannot be subsumed under existing technologies. They are innovative in the sense that they promise new and potentially superior solutions to problems. They are still in development in that they are still largely a promise: no, or not many, products and applications have resulted from them, and few, if any, are marketed and used on a large scale. They are expected to have a large socioeconomic impact in that they are expected to generate significant economic value and activity, and have the promise to affect or transform one or more social or economic domains." (pp. 175-76)

Following this definition, emerging technologies are characterised by high potential: they are still in development but carry the promise of significant impact on what we can do. This impact is sometimes expected to be so far-reaching as to be 'disruptive', urging a recalibration of ethical frameworks (Hopster). While Brey's definition foregrounds the positive dimensions of this high-potential character, emerging

technologies can – and typically do – carry risks and downsides as well. This ambivalence between promise and risk is what makes emerging technology so ethically compelling.

As early-stage, promising technologies with a potentially wide range of applications, quantum technologies are an instance of new and emerging science and technology. The distinction between 'new' and 'emerging' adds nuance here. *New technologies* are technologies that have recently been developed or introduced. They exist and are operational, but are not yet widely diffused or socially embedded. Quantum sensing may qualify in this regard, with several applications operational or in advanced development. *Emerging technologies* are still in the process of taking shape. Existing in a conceptual or early form, their final configuration, capabilities, and applications remain open. Quantum computing and communication typically fall into this category: they exist in early experimental form, but what they will ultimately become remains largely undetermined.

As foreshadowed in the previous section, engaging with the ethics of such technologies comes with a twofold epistemic challenge of novelty and uncertainty. I elaborate on each in the next two sections.

3.2 Understanding: coming to grips with novel technology

Building on Brey's definition, a first defining characteristic of new and emerging technologies is that they confront us with new realities – actual in the case of new technologies, and largely speculative in the case of emerging ones. With their new capabilities they enable new applications, which present us with new situations that give rise to new ethical questions.[4] A first challenge for ethical engagement with new technologies such as quantum is therefore simply to get these new realities into view: to develop sufficient understanding of what a technology entails and what it may do.

Indeed, the emergence of new technologies regularly prompts calls for improving understanding of them – among the general public and among those charged with assessing their impact – to enable informed and grounded engagement. Implicit in these calls is the intuition that meaningful engagement, including ethical reflection, requires some grasp of the technology. This argument also emerged around new quantum technologies (Vermaas, 2017; Coenen & Grunwald, 2017; Seskir et al., 2023; Roberson, 2023). The idea underlying calls for improving the understanding of a technology is

[4] Here, 'new' should not be taken as necessarily unprecedented. It may also include familiar ethical challenges resurfacing in a new context (Swierstra & Rip, 2007; Shelley-Egan & de Jong, 2025). For example, the development of quantum computing may raise questions about dual-use, the phenomenon that a technology developed for civilian or beneficial purposes can also be used for military or harmful purposes, which are questions also raised in the context of other technologies, such as nano technology (Nixdorff, 2018), synthetic biology (Zhang et al., 2024), and artificial intelligence (Vaseashta, 2025).

that in order to engage with it and participate in a discussion about its implications, one needs some degree of understanding of what the technology comprises. While this thought is intuitively plausible, it can only be put to work effectively once we have a clearer account of what such understanding actually entails.

While understanding has been widely discussed in the philosophy of science (Friedman, 1974; Kvanvig, 2003; De Regt, 2017; Reutlinger et al., 2018), it remains underdeveloped in the philosophy of technology. Accounts exist of the knowledge required for the design and use of technology (e.g. Houkes, 2009; Houkes & Meijers, 2022), but a comprehensive account of what it means to *understand* a technology, as a distinct cognitive skill, was lacking. As a result, it remains unclear what it means to understand it sufficiently for a given task – for example, to identify its possible ethical implications.

Without such an account, the role that understanding plays in ethical discussion risks being neglected. Questions may be raised without sufficient grounding in what a technology actually is or can do, and premature or misguided questions may go unrecognised as such. This particularly applies to new and emerging technology, where understanding must still be developed and where much is still unknown. Illuminating what it means to understand a technology and the role such understanding plays in ethics provides a basis for fostering it, and thereby, for ensuring that ethical discussion remains grounded in the realities of the technology under consideration. The first part of this thesis therefore asks:

> *What does understanding mean in relation to technology, and which form of understanding best enables meaningful ethical analysis?*

In the case of new and emerging technologies, understanding is necessarily limited. This points to the second epistemic challenge for ethics of early-stage technology: engaging with uncertainty.

3.3 Timing: engaging in ethics upstream

Alongside its engagement with novel realities, another defining aspect of the ethics of new and emerging technologies is its *timing*. Engagement with early-stage technologies is primarily motivated by the opportunity of early-stage intervention: when technologies are still in flux, the chances of effectively directing their development are highest. In other words, engaging with a technology 'upstream' is an effective way of influencing its consequences 'downstream'. Questions of desirability and risk can be raised before the course of a technology becomes fixed and such questions may become a mere afterthought. The ethics of new and emerging technology is thus, at its core, upstream and anticipatory in nature.

Anticipation, however, comes at a cost. David Collingridge (1980) famously captures this in his "dilemma of social control": at the earliest stages of technological

development, we can effectively influence a technology's trajectory, yet we lack the knowledge needed to identify what we should be steering away from – i.e. an early-stage knowledge problem. By the time we have sufficient insight into a technology's implications, it has already taken a certain shape and redirecting its design and deployment has become considerably harder – i.e. a late-stage power problem. Once the genie is out of the bottle, it is hard to put it back in.

Early-stage ethical engagement pre-empts the power problem by engaging with a technology while it is still taking shape. But this path carries its own difficulty: new and emerging technologies carry the promise of a new world we cannot yet fully see. The earlier we engage, the less we know about what we are engaging with. By anticipating the power problem, we see ourselves confronted with the knowledge problem.

This problem, often referred to as the problem of uncertainty, has been framed as the central problem of ethics of technology (Sollie, 2007). Anyone engaging with the ethics of new and emerging technology has to deal with the uncertainties that surround its subject. Taking the lack of knowledge seriously, some have warned against the pitfalls of what they call 'speculative ethics' (Nordmann, 2007; Rip & Nordmann, 2009). Critiques of speculative ethics, which emerged in the context of nanotechnology in the late 2000s, rejects filling the absence of knowledge about what *is* with speculation about what *may be*. Relying too heavily on hypothetical futures makes ethical analysis shaky. Identifying this means recognising knowledge as a requirement for meaningful ethical debate.

Several approaches have been proposed to cope with the knowledge problem that comes with early-stage or upstream ethics. These include techno-ethical scenarios approach (Swierstra, Stemerding, & Boenink, 2010), the ETICA approach (Stahl & Flick, 2010), anticipatory technology ethics (Brey, 2012), ethical-constructive technology assessment (Kiran et al., 2015), and hermeneutic technology assessment (Grunwald, 2020). These approaches differ in how they handle the epistemic constraints of early-stage ethics. That is, in how they deal with the limited and uncertain information available at a given stage of technological development. This means they are not equally suited to every epistemic situation: some approaches require a degree of knowledge about potential applications or impacts that may simply not yet be available, while others are designed to work precisely under conditions of fundamental uncertainty. Moreover, the knowledge problem is not static: as a technology develops, what is known and unknown shifts, and with it the suitability of different ethical approaches changes.

What remains lacking is a framework for assessing which approach is most suitable at which stage. As a result, the differences between approaches remain insufficiently articulated, and approaches that are appropriate under different epistemic conditions risk being treated as interchangeable alternatives. Applying an approach that requires

knowledge not yet available risks producing shaky or speculative analysis; conversely, failing to move towards more grounded approaches as the technology matures means missing opportunities for more concrete ethical engagement. Calibrating ethical approaches to the epistemic context of the technology is therefore essential for meaningful ethical discussion. This is the focus of the second part of this thesis, asking:

> *How should we calibrate our ethical engagement to what is – and is not yet – understood about a technology?*

While there are approaches in the literature that gesture towards the epistemic requirements for meaningful ethical discussion, it remains unclear what those requirements amount to. Explicit engagements with the epistemic limits of early-stage technology ethics tend to be essentially negative: they warn against speculative ethics and call for keeping ethics in step with technological development rather than outrunning it (cf. Nordmann, 2007; Nordmann & Rip, 2009; Brey, 2012). What is needed is an explicit positive account of what adequate epistemic conditions actually look like. This thesis aims to develop such an account.

4. Chapter overview

In this thesis, I investigate the requirements for a meaningful ethical discussion about technology, especially new and emerging technologies. As outlined in the previous section, developing this account requires addressing a twofold epistemic challenge: gaining sufficient understanding of what a technology is and may do – i.e. the challenge of *novelty and understanding*, and conducting ethical analysis under conditions of fundamental uncertainty about how it will develop – i.e. the challenge of *uncertainty and timing.* The first challenge motivates the first part of this thesis, which develops an account of what understanding amounts to in relation to technology; the second motivates the second part, which examines how the timing of ethical engagement shapes the questions that can meaningfully be asked. Together, the two parts develop several concepts and frameworks that help to navigate these challenges, and put them to work in the emerging ethical discussion about quantum technologies.

What follows is a brief overview of how each chapter contributes to this aim.

Part I: Towards a Theory of Technological Understanding (Chapters 2–4) is a philosophical investigation into what it means to understand a technology. The three chapters develop an account of *technological understanding*, using quantum technologies as a running case to illustrate both the concept and its implications for ethics.

Chapter 2, *A Conception of Technological Understanding*, lays the theoretical groundwork for the concept of technological understanding. It does so by drawing on the discussion of understanding in the adjacent field of the philosophy of science.

Building in particular on De Regt's (2017) account of scientific understanding – the most developed in the field – the chapter develops an analogous account for technology. It shows that in both science and technology, understanding entails a cognitive skill: the ability to successfully use a tool – a scientific theory or a technological artefact – to achieve something. Just as scientific understanding is the ability to explain a phenomenon using a theory, technological understanding is defined as the ability to achieve an aim using a technology. 'Using a technology' is treated here in a deliberately broad sense, covering not only practical use but also designing the technology and envisioning new applications. This breadth points towards different contexts of engaging with technology, each involving its own form of technological understanding. This chapter specifies what technological understanding amounts to in one specific context: the context of design.

The next chapter takes up this thread and spells out what technological understanding amounts to in other typical contexts. **Chapter 3**, *A Contextual Approach to Technological Understanding*, develops the contextual character of technological understanding in full. Complementing the specification for design developed in chapter two, it specifies technological understanding for two further contexts: practical use and innovation. Together, these three contexts capture the main modes of engaging with technology: building it, using it, and envisioning new applications for it. Each context requires a specific type of technological understanding, foregrounding certain aspects of the general notion and backgrounding others. To illuminate what distinguishes the three types of understanding, each is linked to the ability to answer what-if questions at a different level: a technology's inner workings, its direct use-consequences, or its functional capabilities. This pluralistic account provides a basis for clarifying the epistemic requirements for different forms of engaging with technology.

This theoretical work prepares the ground for the question taken up in the next chapter: what is the role of understanding in ethics? **Chapter 4,** *Technological Understanding and Ethics*, draws out the ethical implications of the theoretical account developed in the previous chapters. It presents technological understanding as a requirement for meaningful ethical discussion. More specifically, it argues that understanding a technology's functional capabilities enables tracing a technology's impact. Understanding what the technology is and what it may be used for is a requirement for tracing its possible impact and asking grounded ethical questions about it. This clarifies understanding as an epistemic requirement for meaningful ethical engagement and provides a basis for fostering the groundedness of ethical discussion. It also offers practical guidance: rather than familiarising a wider audience with the physics underpinning a technology – as is frequently done in explanations of quantum technologies for non-experts – efforts to promote public understanding should focus on grasping its potential functions.

Having developed an account of technological understanding and its role in ethics, the thesis turns to a further challenge: how can meaningful ethical discussion proceed when understanding is still limited at the earliest stages of a technology's development?

Part II: Introducing the Ethics Readiness of Technology (Chapters 5-8) is an ethical investigation into how we can engage with early-stage technologies like quantum while taking seriously the uncertainties that surround them. It develops a framework for aligning ethical approaches to a technology's epistemic context, and uses this to engage with ethical questions raised by quantum technologies – about the visions that shape their development, the promises made in their name, and the risks attributed to them.

In order to determine an appropriate approach to the ethics of quantum technology, a framework is needed to assess appropriateness. To this end, **Chapter 5**, *Ethics Readiness of Technology*, develops the concept of *ethics readiness:* the readiness of a technology for specific types of ethical questions. It distinguishes two main types: questions about the outcomes of technological development, and questions about the meanings attributed to it. Linking ethics readiness to the stage of technological development, I argue that the appropriateness of these two approaches shifts as the technology matures and its epistemic context changes. The resulting framework guides which ethical questions are best asked when: an ethics focused on the attribution of meaning is best positioned to deal with the profound uncertainties of the earliest developmental stages, while an ethics focused on outcomes gains relevance as the technology takes shape more concretely.

Chapter 6, *Visions of Quantum Computing*, puts the ethics readiness framework to work. Given that quantum computing is still in an early phase of development, the framework points towards a meaning-oriented approach – one that focuses on the meanings attributed to the technology rather than its concrete outcomes. Visions are a primary vehicle through which such meanings are articulated and circulated: they express how those driving technological development imagine its future and what they expect it to deliver. Since such visions shape the direction of technological development, they are a relevant object of ethical analysis. To investigate them, the chapter develops the method of *ethical vision assessment*. Drawing on an empirical analysis of vision statements by those currently driving the development of quantum computing, four main visions are identified. These visions are scrutinised at the level of their content, rhetoric, and authors, yielding an 'ethical heatmap' of salient issues that warrant further attention.

The next chapter zooms in on one of these flagged issues: the risk of a blind technosolutionist logic. **Chapter 7,** *Technosolutionism and the Case of Quantum for Good,* investigates the phenomenon of Quantum for Good (Q4G) as an instance of technosolutionism. As the normative ambition to mobilise quantum technologies for the benefit of society, Q4G foregrounds the problem-solving abilities of technology.

This echoes a technosolutionist logic that reframes social problems as technological ones. This chapter asks whether that is necessarily a bad thing. Drawing on the literature on technosolutionism, it develops a measured view: technosolutionism can be morally problematic, but is not necessarily so. It identifies the tropes and argumentative patterns that mark the distinction, and assesses the discourse around Q4G against them. It argues that if Q4G is to live up to its commitment – and become more than a normative veneer legitimising strategic innovation agendas – it must be transparent about the limits of quantum solutions for societal challenges and attentive to the sociopolitical and ethical dimensions of the problems it seeks to address.

The final substantive chapter brings multiple lines together. **Chapter 8**, *Risk and Precaution in the Quantum Discourse*, engages with the most pronounced and widely debated implication of a full-scale quantum computer: the unprecedented threat it poses to cyber security (see also Box 1). A thesis about the ethics of quantum technologies cannot sidestep this scenario. At the same time, given the early stage of quantum computing, engaging with a concrete risk of a still speculative scenario sits in tension with the ethics readiness framework (Chapter 5), which normally suggests that outcomes-oriented engagement is premature at this stage. This chapter investigates the conditions under which it is ethically defensible to treat a technology that does not exist yet as a risk demanding precautionary action. To do so, it develops a framework for assessing the plausibility of a risk claim and the proportionality of a precautionary response, and applies it to the quantum cyber security discourse. The analysis largely supports the currently dominant narrative and provides it with explicit ethical grounding. However, it also shows that it rests on implicit assumptions and argues that making these assumptions explicit is necessary for the ethical defensibility of precautionary measures and opens the narrative up to the broader scrutiny it deserves.

Chapter 9, the concluding chapter, synthesises the findings of both parts into an answer to the central research question: *What are the epistemic conditions of meaningful ethical engagement with technology?* It reflects on what the accounts of technological understanding and ethics readiness, taken together, contribute to the ethics of new and emerging technologies, and identifies questions that merit further investigation.

PART I

Part I

Towards a Theory of Technological Understanding

"You only understand things fully when you experience them."

Iain McGilchrist

What does it mean to understand a technology? And why bother? Part I examines the epistemic dimension of engaging with technology. It proposes that such engagement involves a form of *understanding*: the cognitive skill to use a tool to achieve an aim. The following chapters develop an account of such *technological understanding*. Drawing on its scientific counterpart, Chapter 2 articulates a notion of understanding in relation to technology and specifies what this amounts to in the context of design. Chapter 3 introduces two additional contexts – practical use and broader innovation – and argues that technological understanding varies qualitatively depending on the mode of engagement, with each context requiring a distinct type of technological understanding. Recognising this plurality clarifies the epistemic demands associated with different ways of interacting with technology. Building on this, Chapter 4 considers which form of understanding is necessary for ethical reflection, and argues that meaningful ethical engagement depends centrally on understanding what a technology may be used for.

Chapter 2

A Conception of Technological Understanding

An earlier version of this chapter was published as: de Jong, E., De Haro, S. (2026). Technological understanding. On the cognitive skill involved in the design and use of technological artefacts. Synthese, 207(121). DOI: https://doi.org/10.1007/s11229-026-05503-2

Abstract

Although several accounts of scientific understanding exist, the concept of understanding in relation to technology remains underexplored. This chapter addresses this gap by proposing a philosophical account of technological understanding: the type of understanding that is required for and reflected by successfully designing and using technological artefacts. I develop this notion by building on the concept of scientific understanding. Drawing on parallels between science and technology, and specifically between scientific theories and technological artefacts, I extend the idea of scientific understanding into the realm of technology. I argue that, just as scientific understanding involves the ability to explain a phenomenon using a theory, technological understanding involves the ability to use a technological artefact to realise a practical aim. Both theories and artefacts are tools, and using them successfully requires the cognitive skill of understanding. Technological understanding is thus conceived as the ability to recognise how a practical aim can be achieved by using a technological artefact. In a context of design, this general notion of technological understanding is specified as the ability to design an artefact that, by producing a phenomenon through its physical structure, achieves the intended aim. By analogy with De Regt's criterion of the intelligibility of theories, I give a criterion for the intelligibility of a technological artefact, as a precondition for technological understanding. I illustrate the concept of technological understanding through two running examples: magnetic resonance imaging (MRI) and superconducting quantum computers. The account highlights the epistemic dimension of engaging with technology and, by allowing for context-dependent specifications, provides guidance for testing and improving technological understanding in specific contexts.

1. Introduction

What it means to understand a technology has received remarkably little attention. By contrast, in science, 'understanding' has been recognised as a fundamental aim alongside prediction and explanation for several decades, and a number of developed accounts and theories of scientific understanding now exist. The question of what understanding means in relation to technology, however, has remained largely unexplored.

One first line of work addresses the knowledge that is required for both designing and using technological artefacts (e.g. Houkes, 2009; Houkes & Meijers, 2022). However, these accounts do not discuss understanding as the ability to apply such knowledge. In a second line of work, understanding has been discussed in response to 'opaque' technologies[1] and in calls to improve understanding of new technologies to foster public debate (e.g. Vermaas, 2017). Recently, there has also been discourse on whether AI can act as an agent of scientific understanding (Barman et al., 2024; Krenn et al., 2022). In addition, Zhu (2025) calls for a hermeneutic notion of understanding to explore how users make sense of their interactions with technological artefacts and how such artefacts come to have meaning for them.[2] While these contributions valuably add to the emerging discussion, a comprehensive, epistemically informed philosophical account of what it means *to understand a technology* has yet to be developed.

The lack of such an account renders calls for improved understanding of specific technologies vague and potentially ineffective. Furthermore, without a clear explication of what it means to have understanding in relation to a technology, the cognitive dimension of our interactions with technological artefacts remains obscure. This chapter aims to address this gap, by proposing an account of *technological understanding*: the kind of understanding that is involved in designing, reasoning about, and using technological artefacts.

I develop my notion of technological understanding by drawing on Henk de Regt's (2017) account of scientific understanding. In this account, understanding is defined as a cognitive skill that enables one to apply scientific knowledge to perform specific tasks – most notably, explaining phenomena. This requires that the knowledge or theory in question is intelligible to the user. I argue that this conception of understanding offers

[1] For example, the lack of understanding of how certain Artificial Intelligence (AI) models work has been problematised and has given rise to the pursuit of 'explainable AI' (XAI) (Barredo Arrieta et al., 2020; Samek & Müller, 2019; Goodman & Flaxman, 2017).

[2] An important difference between my account and Zhu's call for technological understanding is that, while for Zhu (2025, p. 122) the concept of technological understanding 'refers to the way in which we make sense of how to interact with a technical artefact' (i.e. it takes our interaction with a technical artefact as its primary focus), the object of understanding in my account is a practical aim. I will discuss this further in Section 2.

a valuable framework for developing a similar notion in the realm of technology. Similar to the development and use of theory in science, the design and use of technological artefacts require an epistemic ability[3] to apply knowledge. This involves the cognitive skill of anticipating the consequences of the artefact's operation within a given context.

In Section 2, I first introduce De Regt's notion of scientific understanding. I then give my conception of a technological artefact, and argue that there is an analogy between scientific theories and technological artefacts in terms of the instrumental roles they play in their respective domains.[4] Just as science seeks to understand and explain natural phenomena through the formulation and use of theories, technology seeks to solve practical problems through the design and use of artefacts. I use this analogy as a heuristic to transfer the notion of scientific understanding to the technological domain, thereby developing an analogous – yet distinct – notion of technological understanding.

In Section 3, I extend De Regt's account into the realm of technology, to obtain an analogous notion of technological understanding. De Regt emphasises the pragmatic aspect of understanding, explaining it as the ability to use knowledge. I endorse this view and argue that using a technological artefact similarly involves a type of understanding: namely, it requires the cognitive skill to recognise how the artefact's operation (under varying circumstances) can achieve a desired outcome. Just as scientific understanding entails the ability to use a scientific theory to explain a phenomenon, I argue that technological understanding entails *the ability to use a technological artefact to realise an aim*. The analogy between the two types of understanding thus highlights the cognitive skill involved in using a tool (a scientific theory versus a technological artefact) in achieving a goal (explaining a phenomenon in science versus realising a practical aim in technology). By analogy with De Regt's criterion of the intelligibility of theories, I also give a precondition for technological understanding: namely, a criterion for the intelligibility of a technological artefact.

From the outset, I wish to stress that I use the term 'use' in a broad sense, as employing something (here: a tool) for a purpose. In this conception, 'use' can refer to mental activities (such as employing a theory to explain a phenomenon, or recognising how an artefact's capabilities could serve an aim) as well as practical activities (physically employing an artefact to achieve an aim).[5] This allows me to draw an analogy between

[3] I take 'an ability' to be the potential to perform an action (see also Miller, 2022, p. 470).

[4] I do not present the analogy between scientific theories and technological artefacts – and by extension, between scientific and technological understanding – as evidence in itself, but as a heuristic for conceptual development. The analogy serves to transfer the structure of De Regt's account of scientific understanding into the technological domain, thereby helping us articulate a notion of technological understanding.

[5] Thus I endorse the standard practice in the literature on understanding, explanation, and models, of taking 'use' in the broad English meaning of the term: in the current context of explaining and

using a theory and using a technological artefact, while also leaving room for interpretative flexibility: the ability to use a technological artefact may involve actual practical use, but it can also include thinking up its design or recognising potential applications (see Section 3.2). In other words, how we specify 'using an artefact' affects the scope of technological understanding.

Thus, depending on the specifications of 'use', sub-types of technological understanding can be formulated, which typically apply in certain contexts. In this chapter, I specify the general notion of technological understanding for a specific context: the context of design.[6] To have a design-type of technological understanding entails the ability to realise an aim by *designing* a technological artefact. This type of understanding emphasises the role of knowledge about physical phenomena as well as the cognitive skill that is required to make such knowledge productive to realise a practical aim.

To illustrate this account, I apply it to two examples: magnetic resonance imaging (MRI) and superconducting quantum computers. MRI, a first-generation quantum technology, has been in practical use for decades, while quantum computing, a second-generation quantum technology, has yet to fully materialise. By examining these examples, I illustrate what it means to possess, or to lack, technological understanding.

Giving a detailed account of technological understanding, this chapter clarifies what it means to 'understand' a technology. Specifically, the notion of technological understanding elucidates the epistemic dimensions of designing and using technological artefacts, thus highlighting the cognitive skills that these activities require. In this way, this account acknowledges the active, problem-solving nature of technological practice, where the success of understanding is not just measured by the ability to explain, but by the ability to realise practical outcomes. Furthermore, this account advances beyond existing accounts of understanding by offering a means of specifying it across different technological contexts, thus enabling more precise approaches to fostering understanding of both established and emerging technologies.

---

understanding, it is used to express functionality, instrumentality, and being an epistemic tool. Consider the following examples of various kinds of epistemic use in Frigg (2023): a hammer is used by a doctor to rest reflexes (p. 393), formal tools are used (pp. 26, 53, 96), equipment and physical instruments are used (pp. 82, 84), linguistic terms are used (p. 138), the experimental method is used (p. 90), and data is used as evidence (p. 92). Morrison and Morgan (1999) attest to the same varied usage of the word 'use' in the context of models that are used for epistemic purposes (pp. 11–12, 16, 21, 23, 28). Finally, De Regt (2017) explains how Maxwell and Boltzmann used mechanical models for understanding (pp. 199, 205, 213): indeed, Maxwell's 'use' includes the designing of these mechanical models.

[6] Design is taken as the starting point because it is structurally closest to scientific understanding – on which the account draws – and because it is conceptually prior to other forms of engagement with technology: other forms of engagement with technology presuppose that a technology has been designed.

2. Setting the stage

In order to extend the concept of scientific understanding into the realm of technology, I introduce three ingredients: namely, a conception of scientific understanding (Section 2.1), an account of technological artefacts (Section 2.2), and an analogy between scientific theories and technological artefacts (Section 2.3).

2.1 Scientific understanding

In addition to scientific explanation, scientific understanding has been distinguished as a distinct cognitive ability, different from e.g. knowledge, and as a central aim of science (see for example De Regt & Dieks, 2005; De Regt, 2009; Boon, 2009; Strevens, 2013). According to what Reutlinger et al. (2018, p. 1081) call 'typical accounts of scientific understanding', scientific understanding requires that there is a (true) scientific explanation of a phenomenon, and that this explanation is *epistemically accessible*.[7] In such accounts, a phenomenon P is understood by a scientist S if and only if there is an adequate theoretical explanation for P that is epistemically accessible to S. This raises a key question: What does it mean for an explanation to be epistemically accessible?

De Regt and Dieks (2005), and later De Regt (2009, 2013, 2015, 2017), offer a well-developed theory to address this issue. De Regt defines the epistemic accessibility of an explanation – and thus, scientific understanding – as the ability to use a theory to give an explanation (2005, p. 142). This leads De Regt (2017) to formulate the Criterion for Understanding Phenomena (CUP):

> **CUP**: A phenomenon P is understood scientifically if and only if there is an explanation of P that is based on an intelligible theory T and conforms to the basic epistemic values of empirical adequacy and internal consistency.[8] (2017, p. 92)

[7] There is a debate in the literature about whether scientific understanding is objective, or whether it is a subjective concept, of no interest for the philosophy of science. Thus in an oft-quoted passage, Hempel (1965, p. 413) writes that "such expressions as 'realm of understanding' and 'comprehensible' do not belong to the vocabulary of logic, for they refer to the psychological or pragmatic aspects of explanation". However, Hempel (1965, p. 337) can also be read as holding a more moderate eliminativist view, that 'explanation is understanding enough' (for a summary of this debate, see De Haro and Butterfield (2025, pp. 531–534)). Recently, a number of authors have also distinguished between the use of 'pragmatic' as 'subjective', and the broadly Wittgensteinian sense of 'useable for certain aims'. This sense is compatible with an objective (appropriately inter-subjective) notion of understanding. I here endorse this consensus: see e.g. Friedman (1974, pp. 18–19), Salmon (1978, p. 684), Kitcher (1981, pp. 509, 529; 1981, p. 419), Schutz and Lambert (1994, p. 66), Lipton (2004, p. 30), De Regt and Dieks (2005, p. 150), Grimm (2010, p. 337), and Reutlinger et al. (2018, p. 1081).

[8] The notion of scientific understanding has frequently been discussed in relation to the deductive-nomological (D-N) model of scientific explanation. According to this model, notably advocated by Hempel (1965), a scientific explanation is a deductive argument that derives the occurrence of a

In this criterion, the epistemic accessibility of an explanation is given by the intelligibility of the theory used. De Regt defines intelligibility as 'the value that scientists attribute to the cluster of qualities of a theory … that facilitate the use of a theory' (2017, p. 40). Thus, for scientists to use a theory to explain phenomena, the theory must be intelligible to them. As one way (among others) to objectively assess this, De Regt proposes a Criterion for Intelligible Theories (CIT):[9]

> **CIT**: A scientific theory T (in one or more of its representations) is intelligible for scientists (in context C) if they can recognise qualitatively characteristic consequences of T without performing exact calculations.[10] (2017, p. 102)

Scientific understanding is thus explained as qualitative insight into the consequences of a theory – that is, a skilled recognition of those consequences.

This intelligibility criterion 'captures the pragmatic and contextual nature of intelligibility (and accordingly of understanding)' (De Regt & Dieks, 2005, p. 151), and allows for variation in standards of intelligibility – and by extension, of achieved understanding – among scientific communities. De Regt emphasises that intelligibility is not an intrinsic property of a scientific theory. Rather, it depends on both the characteristics of the theory and on the relevant scientific community that uses it. In other words, the intelligibility of a theory requires a match between its theoretical virtues and the cognitive skills, background knowledge, problem-solving practices, and aims of the scientific community using it. This makes intelligibility a *contextual* notion: it cannot be assessed independently of the context in which a theory is used. This implies that scientific understanding is context-dependent too: if the intelligibility of a theory depends on the user context, then scientific understanding – defined as the ability to use that theory to explain phenomena – also depends on this context.

By emphasising the role of the cognitive and epistemic skills that enable the use of a theory, De Regt's account highlights the *pragmatic* character of scientific understanding. Thus, in his view, understanding is not merely the possession of

---

phenomenon P from general laws of nature in conjunction with specific conditions. De Regt himself adopts a pluralistic stance on scientific explanation, asserting that his account of scientific understanding is compatible with various models, including causal, contextual, and non-deductive approaches. This pluralism acknowledges the complexity of scientific practice and recognises that different contexts may require different explanatory frameworks.

[9] Boon (2009, p. 6) has criticised De Regt's CIT, arguing that intelligibility is not best defined as the ability to recognise qualitatively consequences of T; intelligibility can also be defined more broadly, as the ability to use T in reasoning to solve relevant problems.

[10] This formulation leaves room for variation of standards of intelligibility, and hence, of scientific understanding; for example, across scientific communities. This does not, however, make intelligibility a subjective notion. For, even though its standards can vary, the criterion itself does not change, and it functions as an objective test for theories (see also footnote 7, about the objectivity of scientific understanding).

knowledge, but the ability to actively use knowledge[11] to explain phenomena – it is an 'extra cognitive ingredient' that is required to use a theory successfully (De Regt & Dieks, 2005, p. 149). This involves cognitive skills such as making explanatory connections, answering what-if-things-had-been-different questions, and providing clarifying examples (see also De Regt's latest elaboration in Barman et al., 2024). For example, beyond knowing that there is global warming and that this is mainly caused by the increased atmospheric levels of carbon dioxide, scientific understanding of climate change entails the ability to explain this causal relationship by using an intelligible theory.

This pragmatic character of De Regt's account of scientific understanding, where understanding is defined as the ability to actively use knowledge to perform certain tasks (Barman et al., 2024), makes it particularly relevant to my project of developing a similar notion within the practical and goal-oriented domain of technology (see also Sects. 2.2 and 2.3).[12] Tentatively, it appears that the design and use of a technological artefact also requires an epistemic ability to use knowledge, including a cognitive 'intuition' for the consequences of the artefact's operation, given a certain context. Furthermore, the contextual notion of intelligibility – defined as relative to both the aims of the users and the skills they possess – aligns well with the practice of using technological artefacts. To extend the concept of scientific understanding into the realm of technology, I begin by discussing the technological counterpart of a scientific theory: the technological artefact.

2.2 A definition of technological artefacts

2.2.1 A designed, functional object

Just as scientific theories are usually taken as the core units of analysis in science, I will take technological artefacts to be the primary units of the analysis of technology.[13] Like scientific theories, technological artefacts are constructed or designed objects. Note that what makes artefacts technological is their inherently 'functional', 'useful', and 'instrumental' nature (Kroes, 2002, pp. 292–294; Houkes & Vermaas, 2010; Van de Poel, 2009, p. 980). Hughes even speaks of the 'instrumental function' of artefacts,

[11] In the rest of this chapter, the usage of 'knowledge' in the context of technological understanding will encompass a combination of causal knowledge and factual knowledge of laws, theories, technological and design procedures, and background conditions (see Section 2.1).

[12] As noted in the introduction, Zhu (2025) calls for a hermeneutic notion of understanding in relation to technology that emphasises the way in which we 'make sense' of our interactions with technological facts. My account rather focuses on the epistemic and pragmatic dimension of understanding in relation to technology (see also footnote 2).

[13] I use 'technology' in an abstract sense, not referring to a particular object, and 'technological artefact' as a concrete object with a specific function.

to highlight that they are designed and valued for their effectiveness in performing tasks or solving problems (Hughes, 2009, p. 181). In this thesis, I take technological artefacts to be *designed, functional objects.*[14]

To develop a nuanced notion of understanding in relation to technology, particularly when considering more complex technologies, it is helpful to explore the functional nature of artefacts in greater detail. To this end, I draw on the work of Kroes (2002). Kroes (2002) explains the functionality of technological artefacts through two key aspects: (i) an intentional aspect and (ii) a physical aspect. The intentional aspect is the function attributed[15] to an object, shaped by human intentions. The physical aspect comprises the material properties and structure that make it possible to fulfil this function. I endorse this dual-aspect view, and define a 'technological artefact' as a functional object that possesses both *intentional* and *physical* aspects.

2.2.2 Physical aspects

To deepen the understanding of the physical aspect[16], I break it down into two components: a physical structure and a physical phenomenon:

> *Physical structure*, X: This is the physical construction of the artefact, capable of producing a specific phenomenon as the result of its components working together. It is the material parts – like the metal head of a hammer or the chips in a computer – that make functionality possible.

Since the physical structure consists of parts that work together, it will be useful when analysing a particular artefact to denote these by $x1$, $x2$, etc. For example, in magnetic resonance imaging (MRI), a distinction can be made between the main superconducting magnet that polarises the sample of tissue (i.e. $x1$), the shim coils that correct the shifts in the homogeneity of the magnetic field (i.e. $x2$), the gradient system that is used to localise the region of tissue that is being scanned (i.e. $x3$), etc.

---

[14] Defining technological artefacts as functional objects in terms of realising practical aims does not deny that functional requirements are not the only considerations that shape an artefact (see Van de Poel, 2009, pp. 986–987 about 'additional requirements'), nor does it exclude the related possibility of artefacts having functions that we would call symbolic rather than practical – although a practical function is essential in the conception of technological artefacts developed here.

[15] Kroes (2002) distinguishes a context of design and a context of use, both of which can be constitutive of the technological function. In this chapter, I focus on the intentional context of design and hence on the functionality that is defined and assessed at this stage.

[16] The physical structure includes seemingly non-physical objects, like software, which are realised with various types of physical supports. In this chapter, I focus on the initial or fundamental (that is: material) design process, which explains the focus on the physical aspect of technological artefacts. With appropriate modifications, and using the analogy between technological artefacts and scientific theories, this framework can also be adapted to the study of data-based design processes such as software engineering.

In the case of a superconducting quantum computer, the physical structure consists of several key components. The main superconducting circuits (i.e. $x1$), known as the quantum processing units, house the superconducting qubits ('quantum bits', i.e. quantum analogues of classical bits) or transmons, which themselves have subcomponents like a Josephson junction, a capacitor for external interaction through photon absorption, etc. The control electronics (i.e. $x2$) generate precisely timed microwave pulses that manipulate the states of the qubits, and the cryogenic systems (i.e. $x3$) maintain the ultra-low temperatures necessary for ensuring qubit stability and minimising thermal noise, which can disrupt the fragile quantum states. Other components include coaxial cables, which transmit signals with minimal loss, shielding from external electromagnetic interference, etc.

> *Physical phenomenon*, P: This is the functional phenomenon resulting from the operation of the physical structure. It is the falling weight of the hammer or the electric signals running through computer chips.

In the example of MRI, the phenomenon is the production of the digitalised image of the region of tissue being scanned. As a *phenomenon*, this image is an uninterpreted material object, produced by a series of physical and chemical operations that follow the MRI's operations. Like with the physical structure, it is possible to distinguish several constitutive sub-phenomena within the phenomenon P, like the creation of the magnetic field with particular properties of homogeneity (i.e. $p1$), the polarisation of the protons in the sample of tissue (i.e. $p2$), the resulting magnetic resonance signal (i.e. $p3$), etc.

In the case of a superconducting quantum computer, the central phenomenon is the behaviour of qubits, which are realised as superconducting circuits or transmons. This behaviour can be analysed into several phenomena. One such phenomenon is quantum tunnelling (i.e. $p1$) where pairs of electrons (Cooper pairs) tunnel through a Josephson junction. This enables the formation of unevenly spaced discrete energy levels (i.e. $p2$), thus allowing the selection of the lowest two states as qubit states, namely the ground state and the first excited state. Another is superposition (i.e. $p3$), where each transmon qubit is in a combination of its ground and excited states simultaneously, enabled by the controlled oscillations in the superconducting circuit. Entanglement (i.e. $p4$) also plays a vital role, where multiple qubits become correlated through shared electromagnetic interactions, allowing for the complex interactions necessary for quantum computation. These are just a few key phenomena; others include quantum measurement, (de)coherence, quantum error correction, etc.

2.2.3 Intentional aspects

The *intentional* aspect of a technological artefact refers to its relation to a specific aim: the technological artefact t realises, achieves, or embodies its intended aim A (see also

Weber et al., 2013). For a technological artefact to realise an aim typically means to solve a practical problem or fulfilling a specific need of a group of (prospective) users. This can be virtually any type of practical need: from personal transportation to communication over distance, and from solving computational problems to providing heat and light. I refer to the aims of technological artefacts as 'practical', since they require the artefact's practical use.[17]

I further break down the intentional aspect into two levels: *direct aim* and *ultimate aim*:

> *Ultimate aim*, A: This is the ultimate effect that the artefact is intended to achieve.[18] The ultimate aim is 'external' to the artefact, in the sense that it can exist independently from the specific artefact, and a given technological artefact can accommodate for various ultimate aims (see also Kroes, 2002).

In the example of MRI, the ultimate aim is to perform a medical examination of organs, tissues, and the skeletal system in a non-invasive way. For a quantum computer, the overarching ultimate aim is to solve a specific class of computational problems that are (practically) intractable for classical computers. Achieving this aim potentially enables a wide range of applications, from accelerating drug discovery and the development of new materials, to optimising logistics and simulating complex natural systems.

> *Direct aim*, a: This is the specific way the technological artefact achieves – or approximates – its ultimate aim through its operation. It is the specific technological response to the ultimate aim – for example, like performing an MRI scan to detect tissue abnormalities is one way to (technologically) support medical diagnosis. The direct aim is implicit in the production of the phenomenon; they are two sides of the same coin. The direct aim adds a layer of interpretation of the produced phenomenon in terms of intentionality. I make this explicit in my notion by writing the technological artefact t as a triple: $\langle \mathsf{X}, \mathsf{P}, \mathsf{a} \rangle$.

The ultimate aim of the MRI can be technologically realised by the MRI's direct aim of producing medical images of the inner human body according to set specifications, i.e. images that are interpreted by a doctor. In the case of the quantum computer, the ultimate aim of unlocking increased computational abilities can be realised through

[17] By characterising the aim of a technological artefact as practical, I do not wish to contrast 'practical' and 'theoretical'; since in the previous section I already said that artefacts can be used for epistemic aims (think of e.g. telescopes and particle accelerators). Rather, 'practical' here indicates that the realisation of the aim requires the artefact to do something, i.e. to perform a task that leaves the artefact and its immediate environment in a different state that it was before.

[18] A technological artefact can – and often will – have other effects and uses than initially foreseen or intended during the design process. In this chapter, I limit the scope to the intended use of an artefact and its associated effect.

the direct aim of executing specific quantum algorithms on a fault-tolerant quantum computer – that is, a quantum computer equipped with a sufficient number of reliable qubits that possess adequate coherence times, dependable quantum gate operations, robust error correction mechanisms, and high measurement fidelity.

The distinction between the artefact's direct and ultimate aim is useful for two reasons. First, it allows us to distinguish between the aim as it is instantiated by the artefact (namely the direct aim, e.g. generating a medical image with a specific resolution quality), and the aim that can be conceptualised independently of the specific artefact (the ultimate aim, e.g. achieving improved medical diagnosis). The distinction between the ultimate aim A and its embodiment a in a technological artefact, is analogous to the distinction between a phenomenon P and the description of that phenomenon given by a scientific theory. A scientific theory provides a particular description or perspective on P; in a similar way, the direct aim articulates a particular interpretation of the ultimate aim A. Both a theory and an artefact's direct aim thus function as mediating descriptions that connect a broader object (a phenomenon or an ultimate aim) to a concrete way of engaging with it.

Second, it highlights that a is often an approximation of A. As Van de Poel (2009, p. 985) emphasises, an overview of the complete set of possible solutions to a design problem (i.e. the problem for which the artefact-to-be-designed is supposed to offer a solution) is usually lacking. The aim a captures only one way of operationalising the ultimate aim A, and could even do so partially. In this sense, a represents a solution to A that is 'good enough' given the circumstances. I use the term 'approximation' to emphasise the open-endedness of specifying A in practice: each direct aim reflects a context-dependent – and often partial – way of realising the ultimate aim.

As was already pointed out, it will be clearest to consider the ultimate aim as *external* to the artefact. That means that the ultimate aim A is, strictly speaking, not part of the technological artefact t, i.e. the triple $\langle X, P, a \rangle$. While a technological artefact t requires *some* ultimate aim A to be considered functional, it does not need to be tied to a *specific* ultimate aim from the outset. Furthermore, the ultimate aim of an artefact can change over time. Although one could argue that (the meaning or character of) an artefact changes depending on what it is used for, I maintain that the artefact itself is the same for various ultimate aims (so long as X, P and a do not change). Thus fixing or determining a *specific* ultimate A is not required to define a technological artefact. In this way, the technological artefact and the ultimate aim are logically independent, and they are related by the degree to which, once fixed, the direct aim matches the ultimate aim. Therefore, the question of the degree to which a technological artefact t serves an ultimate aim A is a normative question about how well the direct aim a realises or approximates the ultimate aim A.

2.2.4 Definition of a technological artefact

My conception of a technological artefact, i.e. $t = \langle \mathsf{X}, \mathsf{P}, \mathsf{a} \rangle$, is summarised as follows:

> A **technological artefact** is a designed object, consisting of a physical structure X that produces a physical phenomenon P to achieve a direct aim a, which serves an ultimate aim A

This 'assembled' conception of technological artefacts – defining it as an interplay of physical phenomena, physical structure, and intended functionality – is embedded in extant accounts by Kroes (2002), and Houkes and Vermaas (2010), and Boon's work on the role of physical phenomena in the engineering sciences (2017). The aim here is not to present a wholly novel account of artefacts, but to foreground and integrate specific aspects into one, multifaceted definition. By explicitly including the role of phenomena in attaining an aim, and distinguishing between direct and ultimate aims, this conception allows us to demonstrate how understanding in different contexts may focus on different components of the artefact. Additionally, this conception underscores how scientific understanding underpins technological artefacts – an emphasis particularly appropriate for advanced technologies that rely heavily on the production and manipulation of certain physical phenomena, such as the polarisation of nuclear spin or quantum entanglement which are at the core of quantum technologies.

In the next section, I use this conception of technological artefacts as a conceptual building block to develop an analogy between science and technology and, more specifically, between scientific theories and technological artefacts.

2.3 An analogy between scientific theories and technological artefacts

To extend the concept of understanding from the scientific realm into the realm of technology, I must first explore the relationship between these two fields.

Science and technology have a 'symbiotic' relationship (Barnes, 1982, p. 168), in which they inform and shape one another. Just as scientific breakthroughs lead to the development of new technologies, technological advancements also often enable new scientific discoveries (Boon, 2015). Thus technological artefacts can be seen both as outcomes of scientific understanding and as contributors to its advancement.

This mutual relationship is grounded in a shared concern with understanding physical phenomena. In the uses of both a theory and an artefact, a phenomenon is 'generated' as an outcome: respectively conceptually through theoretical derivation, explanation and prediction, and practically through the operation of the artefact. Furthermore, in experimental science, these two processes go together when phenomena are practically produced in the lab.

The aims of their respective concerns, however, differ (Boon, 2006, 2009). In science, the concern with phenomena is first and foremost *epistemic* (De Regt, 2009): the primary goal is to explain and understand phenomena, with scientific theories and models serving as our tools for doing so. In technology, the concern with understanding phenomena is *instrumental*: the goal is to 'exploit' phenomena (i.e. make them productive) to achieve practical aims, using technological artefacts as tools (see also Boon, 2020: pp. 80–94). However, these aims can be considered deeply intertwined: explanatory information is information that is (potentially) relevant to manipulation and control, and vice versa (Woodward, 2003).

This perspective highlights that the technological exploitation of phenomena is always based on an underlying explanation, whether implicit or explicit (see also Woodward, 2003; Boon, 2006, p. 27). In this sense, technology invariably builds on a – contextual, and more mundane or more advanced[19] – scientific understanding of phenomena, somewhere down the line. For example, designing a laptop relies on scientific understanding of the semiconducting properties of transistors in processing chips. Similarly, designing a quantum computer depends on an advanced grasp of quantum phenomena like superposition and entanglement. Every technological artefact uses and takes advantage of scientific understanding of phenomena, making it productive for some practical aim.

Technological artefacts can thus be considered (physical) instantiations of scientific knowledge and explanations. In that respect, artefacts can be seen as analogous to scientific theories: where a scientific theory can be used to explain a phenomenon abstractly, a technological artefact operationalises the explanation[20] in a concrete, material form, thus 'embodying' the theory.[21] The use of a technological artefact can be thought of as a special instance of using scientific knowledge, which is why the concept of scientific understanding (as the ability to use a theory) offers an interesting starting point for conceiving a notion of understanding in the context of technology (Boon, 2015).

The analogy between scientific theories and technological artefacts thus centres on their engagement with physical phenomena. Scientific theories and technological artefacts are analogous in that both function as tools bridging physical phenomena and human aims: theories provide explanations of phenomena (to achieve understanding), while artefacts exploit these phenomena (to realise practical outcomes).

---

[19] The degree to which a technological artefact builds on scientific understanding depends on the complexity of the technological artefact.

[20] While there are various models of explanation relevant in science, the ones that are most relevant to the design and use of technological artefacts are the causal and the mechanistic models (see also Boon, 2006, p. 27).

[21] For a recent account of the usages of 'theory' and 'model', see Frigg (2023).

Similarly, Morgan and Morrison (1999) argue that there is an analogy between (theoretical) models and technology. They emphasise that models function as tools or instruments, that facilitate recognising how a theory is applied in specific circumstances (pp. 10–11). In that sense, they argue, a model holds the quality of a technology, demonstrating its power in its use. Furthermore, they stress that models can work as useful instruments for the design of technologies since 'they provide the kind of information that allows us to intervene in the world.' (p. 23) While artefacts can indeed be perceived as akin to models in their capacity to bridge the gap between theory and practice, I advocate for a more direct analogy between scientific theories and technological artefacts, endorsing the view that theories serve as foundations of models, and that artefacts typically operationalise general theoretical principles (see, in particular, De Regt, 2017, pp. 31–36).

However, this is not to say that scientific understanding is sufficient for the design of technological artefacts (cf. Zhu, 2025). As Boon (2006) notes, 'using science' in artefact design requires 'substantial additional work.' This additional work relates to the nature of technological artefacts, which I previously defined as objects that are capable of producing specific phenomena through a physical structure to achieve a particular aim. Understanding the functional phenomenon is only one aspect of the process; other aspects include choosing an appropriate phenomenon given the particular aim, and thinking up a physical structure that can produce that phenomenon. This distinction hints at the differences between theories and artefacts, which I will explore further in the following section.

Given this shared focus on understanding physical phenomena, and the way technology builds on scientific explanations to exploit those phenomena, I thus propose that there is an analogy between science and technology, particularly between scientific theories and technological artefacts. While there are important similarities between theories and artefacts, there are also key differences – hence the analogy, rather than identity. The analogy primarily serves to validate the relevance of employing the concept of scientific understanding in developing a notion of understanding in the context of technology. As discussed in Section 2.1, this relevance predominantly stems from the notion of scientific understanding as a cognitive skill to perform certain actions. Ultimately, the notion of technological understanding is intended to stand on its own, independent of the analogy used to construct it. Its usefulness will be illustrated by how it clarifies what it means to understand technology in different contexts.

3. A conception of technological understanding

In this Section, I develop a conception of technological understanding, in analogy with scientific understanding. I first give a general conception of technological understanding

(Section 3.1), followed by a discussion of the scope of this concept (Section 3.2), and a specification of technological understanding for a context of design (Section 3.3).

3.1 Technological understanding: the ability to use a technological artefact

As already discussed in Section 2.3, there is an analogy between scientific theories and technological artefacts. For example, both serve as an instrument to achieve specific aims: scientific theories explain phenomena, i.e. answering the question *Why (does) P (occur)?*, while technological artefacts are designed to realise practical aims, i.e. solving a problem or meeting a need. This analogy was strengthened by the idea that a technological artefact operationalises the explanation of a phenomenon, making it productive in its operation. By doing so, the artefact 'exploits' the phenomenon to achieve its intended function.

I use this analogy between scientific theories and technological artefacts as a stepping stone for transferring the notion of scientific understanding to the context of technology.[22] If we think of understanding as the cognitive skill to (recognise how to) use an instrument to achieve an aim in varying circumstances, then successfully using a technological artefact indeed involves a type of understanding. Building on the concept of scientific understanding as the ability to use a theory to explain a phenomenon, I propose *technological understanding* analogously as the ability to use a technological artefact to realise a practical aim.

It is important to note that 'the ability to use a technological artefact to realise an aim' is the conception of technological understanding in its most generic form: the 'use' of a technological artefact is understood here broadly, and includes design and (conceptual as well as actual) use. This broadness mirrors De Regt's notion of scientific understanding, which applies in both the context of developing a new theory, and in the context of using an existing theory to explain a particular phenomenon. In the next section, I explore in more detail how this general conception of 'use' can be scoped and refined. For now, however, I leave 'use' open to further specification.

In this generic sense, I describe the ability to use a technological artefact as the capability to provide an appropriate 'technological response' to a specific aim – one that successfully achieves the desired outcome. This may involve designing, operating, or conceptualising the artefact (see Section 3.2), and it requires insight into the need or demand itself as well as into how it can be technologically addressed. Thus technological understanding is the cognitive skill that is both demonstrated in, and necessary for, the successful use of a technological artefact.

[22] In this sense, the concept of technological understanding should ultimately stand on its own merits, independently of the analogy.

This cognitive skill can be explained as an ability to use knowledge of the artefact's operation flexibly across different contexts. In other words, having technological understanding means being able to reason about how an artefact can be used to achieve a given aim. Like scientific understanding, technological understanding involves explanatory reasoning. However, while scientific understanding does not privilege a particular dimension of explanation, technological understanding stresses the causal dimensions of explanation. Because its purpose is to bring about a particular state of affairs, technological understanding requires recognising how the artefact's operation leads to the realisation of the intended outcome.

One demonstrates technological understanding when one can foresee how an aim can be achieved through the use of a technological artefact: for example, understanding how an umbrella can be used to stay dry during a rain shower while anticipating that strong wind may reduce its effectiveness. In the case of an MRI machine, technological understanding involves recognising how it can be used for non-invasive medical examinations. Similarly, for a superconducting quantum computer, technological understanding would involve the ability to recognise how to use it for performing the kind of computations that, for example, could accelerate drug discovery.

Defining technological understanding as the ability to realise an aim by using a technological artefact implies that the object of technological understanding is the ultimate aim A. Having technological understanding of an aim means recognising how it can be achieved by the operation of a technological artefact (i.e. t). Thus the matter to be technologically understood is not the artefact itself but a desired outcome – just like it is the phenomenon that is the object of scientific understanding and not the theory itself. However, since technological understanding involves recognising how the artefact's operation realises an aim, it requires a level of 'understanding' of the artefact itself. In other words, technological understanding requires that the technological artefact is *intelligible*. While technological understanding is, strictly speaking (a specific kind of) understanding of an aim, it could also be informally described as 'understanding a technology'.

Building on De Regt's CUP for scientific understanding (see Section 2.1), I propose a Criterion for Technological Understanding (CTU):

> **CTU**: An aim A is technologically understood if it can be realised by using a technological artefact t

Applying this criterion to the MRI example, it becomes clear that technological understanding is achieved: the aim of non-invasive medical examination is realised through the use of the MRI machine. By contrast, there is not yet full technological understanding of the aim to accelerate drug discovery through a quantum

computational advantage,[23] since a fully functioning quantum computer does not yet exist to achieve this aim.

To achieve scientific understanding of a phenomenon P, one must select or develop a theory that is capable of adequately explaining it. Similarly, technological understanding of an aim requires selecting an artefact that can successfully achieve the desired outcome – or, if such an artefact is not available, designing it (see Section 3.3). Selecting an adequate artefact is implied by having technological understanding.

Since the ability to use a technological artefact requires that the artefact is intelligible to the user, I also propose a Criterion for Intelligible Technological Artefacts (CITA), inspired by De Regt's CIT:[24]

> **CITA**: A technological artefact t is intelligible for a subject S (in context C) if S can recognise qualitatively characteristic consequences of t's operations without practically performing these operations

The ability 'to recognise qualitatively characteristic consequences' can be explained in terms of the ability to explore the space of possible initial and end-states of the artefact without practically performing its operations (and regardless of whether, and how, the ultimate aim A is realised). Such exploration requires cognitive skills and reflects the epistemic activity that is involved in technological understanding. This can be characterised as a qualitative recognition of the consequences of the artefact: a skilled judgement of how it works and how it might be used.

CITA's explicit reference to the subject[25] and the context points to the pragmatic and contextual character of the artefact's intelligibility, and so, of technological understanding. The intelligibility of an artefact cannot be established in isolation, but only in reference to a (typical) subject: the same artefact might be intelligible to one (type of) user, but a black box to another. Intelligibility thus depends on the match between the artefact and contextual factors such as skills, familiarity, and the role the user: whether the artefacts is deemed (sufficiently) intelligible can only be assessed in reference to the subject *to whom* it is (un)intelligible. In other words, intelligibility is

[23] A 'quantum advantage' is achieved when the increased computational power enables quantum computers to solve mathematical problems that classical computers cannot solve in practice (Hoofnagle & Garfinkel, 2022).

[24] To flag that there may be other ways to determine the intelligibility of a theory, De Regt (2017, p. 102) adds the subscript CIT_1 to his criterion, and explains: 'CIT_1 is a sufficient condition for intelligibility, not a necessary one; there may be alternative criteria CIT_2, CIT_3, etc.' Although, for simplicity of the notation, I will not use the subscript in this chapter, I am equally open to the possibility of testing the intelligibility of a technological artefact in other ways.

[25] Our subject stands for a typical subject of the relevant community, and is thus contextually determined.

contextual and allows for variation among different groups of users, which in turn leads to variation in technological understanding.[26]

To illustrate, consider the MRI machine. A medical technician or radiologist understands the MRI's operation well enough to anticipate how adjustments in magnetic fields and radiofrequency pulses will produce different types of images, such as distinguishing between soft tissues or identifying specific abnormalities. Even without running every possible scan, they can predict the consequences of various settings and inputs. This ability to foresee outcomes, based on knowledge of the machine's functioning, makes the MRI intelligible to them; lacking this ability renders the MRI unintelligible to many others.

In contrast, with the superconducting quantum computer, while there is a basic understanding of its principles (see Section 2.2), there is not yet a fully intelligible artefact, simply because the design has not yet been successful. There may be scientific understanding of the physics underpinning quantum computing (De Regt & Dieks, 2005; De Regt, 2017),[27] but the lack of practical implementation in a large-scale, fault-tolerant quantum computer with reliable outcomes means that in this case the artefact's operation and its consequences are not yet recognisable. I delve further into this in Section 3.3, where I explore technological understanding within a design context. For now, it is important to see that the absence of an intelligible artefact implies a lack of technological understanding in the case of superconducting quantum computers.

The concept of technological understanding illuminates the cognitive dimension of designing and using technological artefacts. By highlighting that these activities involve cognitive skills and reasoning with imperfect knowledge, it shows that the use and design of technological artefacts require the ability to effectively navigate and use knowledge. Furthermore, technological understanding itself emerges as an epistemic aim, critical to the successful use of artefacts. From this perspective, the successes,

[26] This parallels De Regt's pragmatic, context-dependent conception of scientific understanding, which likewise allows for differences in standards of intelligibility of theories, and hence in scientific understanding.

[27] Indeed, the debate between Heisenberg and Schrödinger over the *Anschaulichkeit* of quantum theory is one of the major historical case studies to which De Regt's theory applies (see De Regt, 2017, Chapter 7). I endorse the clear verdict of the case studies that quantum phenomena, including those like superposition that play a role in quantum computation, can be scientifically understood (to a satisfactory degree) by using quantum mechanics. This verdict disagrees with those conceptions of understanding that require additional conditions (for example, visualisability, or a specific metaphysics based on classical theories, etc.), and that are arguably only reactionary attempts to retain a notion of understanding that is moulded on specific superseded theories. For when it comes to understanding quantum phenomena, the more openminded positions in the neighbourhood of Pauli's and Heisenberg's views, rather than Schrödinger's view of understanding as *Anschaulichkeit*, are vindicated. For a discussion, see De Haro and Butterfield (2025, pp. 119–120) and De Haro and De Regt (2020, p. 3122).

challenges, and failures in designing and using artefacts can be seen as cognitive achievements or limitations, offering a richer framework for analysing technological practice.

3.2 Specifying technological understanding: three logical options

Having outlined a general conception of technological understanding, I now turn to exploring the logical options for its further specification. As mentioned in the previous section, the phrase 'ability to use a technological artefact' has so far remained vague, particularly regarding the meaning of 'use': Does it include design, and does it imply the actual operation of the artefact? In this section, I explore three key options for specifying technological understanding. Doing so will eventually help to clarify what technological understanding entails for typical groups of users, representative of a broader category of subjects.

The available options are based on whether the aim and the artefact are determined, since this affects the cognitive requirements of technological understanding and, ultimately, how 'use' is specified. Depending on whether the aim and the artefact are predetermined (i.e. 'fixed') or underdetermined (i.e. 'open'), I identify three logical options for an account of technological understanding:

(i) *Maximal account*: where the aim A and the technological artefact t are open
(iii) *Minimal account*: where the aim A and the technological artefact t are fixed
(iii) *Via media*: where the aim A is fixed, and the technological artefact t is open; or where the aim A is open and the technological artefact t is fixed

The logical strength, or scope, of these accounts differs with respect to what technological understanding encompasses. In the *maximal account* (i), technological understanding involves both selecting or formulating the aim A and devising a technological artefact t that achieves this aim. I refer to this as the 'maximal' account because it involves the greatest degree of cognitive freedom. This specification of 'use' typically applies to situations where a new connection between an aim and a (prospective) artefact is invented or devised. In this option, the use of an artefact can thus be specified as 'conceptualising' the artefact.

In the *minimal account* (ii), technological understanding does not involve selecting either the aim or the artefact. Instead, it involves enacting an already available connection between the aim and the artefact. This applies to the common employment of the artefact. In this case, 'use' is specified as actual operation, but also in the form of the ability to give an explanation of how the artefact can be operated to achieve the aim. In other words, for this type of technological understanding, one requires an 'explanation of the output from the input and the aims' by reasoning – either explicitly or implicitly by successfully using the technology.

The *via media* (iii) is the intermediate option between the maximal and the minimal accounts. Here, technological understanding involves determining or 'thinking up' the artefact to achieve a given aim, or finding a new aim for an already available artefact. The first scenario typically applies to situations where an artefact is designed to meet a specific, predetermined aim. In this option, 'use' is specified as 'design for use'.

I consider the second scenario (given artefact, open aim) to be a variant of the maximal account (i). In that account, the emphasis is on thinking up the relation of functionality between an artefact and an aim – which closely resembles the process of instrumentalising an existing artefact for a new aim in this second scenario of the via media. In other words, 'repurposing' an artefact can be considered an instance of 'conceptualising' it. Therefore, I do not consider this scenario as a separate option.

By contrast, I treat the scenario of 'open artefact, given aim' as a distinct option because it directly engages with the practice of design. Here, the artefact must be conceived to meet a given aim, which, as I will argue in Section 3.3, requires a unique form of technological understanding that emphasises cognitive freedom in constituting the artefact.

Thus, the three logical options for specifying 'the ability to use an artefact to realise an aim' yield three corresponding specifications or subtypes of technological understanding. Each of these subtypes assumes that the artefact is intelligible to its user (understood here as a representative group of subjects): it is this intelligibility that makes technological understanding possible in the first place. Yet, just as standards of intelligibility for theories may differ across scientific communities, the intelligibility of a technological artefact can vary across user groups.

Each subtype of technological understanding is characteristically associated with a particular context of engagement or 'use', corresponding to the logical distinctions introduced above. These may be described as contexts of design (conceiving or constructing the artefact), practical use (operating it), and innovation (conceptualising new applications for it). This context-sensitivity of understanding is also noted by Zhu (2025, p. 120). I elaborate on this aspect in the next chapter, in which I further explore the three typical contexts outlined here. In the present chapter, I focus on one of these subtypes – the *via media* – which concerns technological understanding in the design context.

Before proceeding, it is important to emphasise that these accounts, along with the specifications of technological understanding they represent, are *logical options*. In practice, these options often overlap and cannot always be neatly distinguished. For example, the process of thinking up a new relationship between an aim and an artefact can include the process of thinking up the artefact itself. In this scenario, a context of (what I will call) 'innovation' overlaps with a context of 'design'. Similarly, in practice, artefacts can be used for other ends than originally intended, so that a context of

'operation' becomes an instance of 'innovation'. However, for conceptual clarity, it is best to distinguish between different scopes – and thus, subtypes of – technological understanding.

3.3 Technological understanding in a context of design

Since 'using a technological artefact' presumes the existence of the artefact itself, it is relevant first to consider a scenario where the artefact has yet to be developed – a situation that mirrors the need to develop a scientific theory in order to achieve scientific understanding.[28] This scenario typically corresponds to a design context.

Kroes (2002, pp. 298-99) characterises the designer's task as turning a description of the intentional aspect of a technological artefact ('functional description') into a description of the physical aspect of the artefact ('structural description'). In my account, the 'functional description' of the artefact(-to-be) corresponds – to some degree[29] – to what I have called the ultimate aim A, while the 'structural description' matches the triple $\langle X, P, a \rangle$ (see Section 2.2). The design process thus starts with a description of the ultimate aim,[30] and succeeds if an artefact can be designed that is capable of realising that aim. In other words, turning a functional description into a structural description involves designing a physical structure X that generates a phenomenon P capable of achieving a direct aim a, which realises the ultimate aim A. This can be reformulated by saying that achieving technological understanding in a design context is a matter of finding an appropriate $t = \langle X, P, a \rangle$ that can realise a specific A (i.e. solving a problem or fulfilling a need).

A key element of this process is the creative step of selecting a phenomenon that can be harnessed to achieve a specific aim. This reveals an asymmetry between scientific and technological understanding. In the process of scientific understanding, the theory-independent question *Why P?* is the starting point. The phenomenon to be explained

---

[28] This is foreshadowed in De Regt's work because his case studies and his own development of the notion of scientific understanding focus on the understanding that scientists have rather than on understanding by non-scientists (even though his own notion in principle applies to both). (Indeed, his CIT explicitly mentions scientists.) Thus I argue that it is also appropriate here to begin this account of technological understanding by focusing on the technological understanding that designers have, and in a second step on the understanding by non-designers. I take this second step in Chapter 3, where I further develop the general notion of technological understanding, CTU, in contexts other than the design context.

[29] The intentional aspect that is described in the functional description is, strictly speaking, part of the technological artefact in the sense that the direct aim a is inherent to the artefact. However, since the ultimate aim A is external to the technological artefact itself, a description of A does not fully correspond to the functional description of an artefact.

[30] As discussed in Sects. 2.2 and 3.2, we consider the ultimate aim to be external to the artefact, and consider formulating it not as a necessary part of designing; rather, design is concerned with shaping an appropriate artefact, and thus formulating the direct aim that adequately achieves the ultimate aim.

is not selected in the process of understanding by theory but is given and fixed; scientific understanding is always relative to a predetermined P. By contrast, in technological understanding, the phenomenon P serves as a means to an end and can be altered if it fails to meet the technological aim (see also Boon, 2015).

The process of identifying an appropriate phenomenon and designing a physical structure to produce that phenomenon requires cognitive skills. Typically, the 'space of possibilities' – the range of all potential technological artefacts capable of achieving the aim – is not fully known. Thus, design is not merely a selection from predefined options: it requires qualitative strategies and judgement, i.e. *an ability to use knowledge*, to effectively integrate X, P, and a. This capacity for reasoning is what I call 'understanding', as requiring more than 'knowledge' alone.

In artefactual design, this reasoning involves causal explanation of how the phenomenon is produced by the physical structure. In other words, to be able to produce, control, and manipulate the phenomenon, one must have sufficient knowledge of how it is caused. Hence, the design process inherently builds upon scientific understanding of the relevant phenomena, and more complex or difficult-to-achieve phenomena demand more advanced scientific understanding. In this sense, scientific understanding is – implicitly or explicitly – part of technological understanding.[31]

In the example of quantum computing, the challenge of selecting an appropriate phenomenon P is illustrated by the ongoing exploration of different approaches to quantum computer design, each focusing on distinct quantum mechanical phenomena (National Academies of Sciences, Engineering, and Medicine, 2019, pp. 127–30). Despite these efforts, none of the current methods has succeeded in creating a sufficiently well-functioning network of qubits that yields significantly increased computational power. This suggests that scientists and engineers have not yet succeeded in identifying and producing a phenomenon P, through a structure X, that successfully instantiates A. For example, for the superconducting quantum computer it remains a challenge how to measure the state of the qubits quickly, below the decoherence time, but still precisely and with high fidelity. In turn, this may be because the right physical structure X, phenomenon P, and conditions to achieve the direct aim a of balancing coherence and measurement fidelity in a way that allows scaling, have not yet been found. In other words, the right kind of triple $\langle X, P, a \rangle$ remains elusive because P and X are not yet intelligible in this context.

Viewing the development of quantum computers through the lens of technological understanding frames this challenge as an epistemic problem of incomplete technological understanding. The central issue in designing quantum computers is not

[31] This is consistent with the work of Boon, who also highlights the role of understanding physical phenomena such that 'instrument […] can be built, controlled, created, calibrated and/or otherwise manipulated.' (Boon, 2015, p. 73; 2017, p. 2).

the scientific question of *Why P?* but the technological one of *Which X for which P for a?* Answering this question is essential for designing the artefact t. This example underscores how selecting a suitable phenomenon in a design context requires more than knowledge: it requires technological understanding to integrate X, P, and a effectively, such that A is achieved.

In summary: Technological understanding is demonstrated in recognising how an aim can be realised by using a technological artefact. This ability can be specified as the ability to design an appropriate technological artefact. The cognitive skill involved in selecting a phenomenon P from among possible alternatives, and designing a physical structure X that can generate P to fulfil the direct aim a, which in turn embodies the ultimate aim A, characterises technological understanding in a design context. Thus successful artefactual design always demonstrates technological understanding, though technological understanding itself is not limited to design (since alternative specifications for 'using a technological artefact' are possible).

Accordingly, the Criterion for Technological Understanding (CTU) in a design context can be formulated as follows:

> **CTU$_{\text{design}}$**: An aim A is technologically understood if it can be realised by designing a technological artefact t, i.e. by constructing a physical object X that, by producing a phenomenon P through a physical object X, achieves a direct aim a that embodies the ultimate aim A

In a design context too, technological understanding hinges on the intelligibility of the artefact. More precisely, the ability to successfully design a technological artefact implies that it must be intelligible. As discussed in the sections on scientific and technological understanding (Sections 2.1 and 3.1), intelligibility is relative to the (typical) subject and depends on contextual factors. In a design context, this means that the artefact must be intelligible to the designer in terms of how its physical structure X and the phenomenon it produces P are capable of achieving the desired aim a, embodying the ultimate aim A. In other words, the success of the design depends on the designer's ability to establish these relations. Therefore, the necessary level of intelligibility is tied to the designer's capacity to reason about how changes in the artefact's physical structure will lead to corresponding changes in the phenomenon produced, and ultimately, in the fulfilment of the aim. Thus, intelligibility of the artefact in a design context is about understanding the qualitative relationship between X and P, and how that relationship can be controlled or manipulated to achieve the intended result.

Table 1 summarises the account of technological understanding up to this point. Rows 1 and 2 capture the analogy between (the function of) a scientific theory and a technological artefact, as discussed in Section 2.3. Building on this analogy, rows 3 and

4 draw comparisons between the types of understanding required for, and demonstrated by, the skilled use of these tools. Row 4a details that the use of a theory or an artefact implies their selection or, where necessary, their development or design. Row 4b highlights that the design of an artefact requires the additional step of selecting a suitable phenomenon, a step that has no equivalent in scientific understanding.

Table 1. Analogy between scientific theories and technological artefacts, and by extension, between scientific understanding and technological understanding

	Unit of analysis	*Scientific theory* $\boldsymbol{T}$	*Technological artefact* $\boldsymbol{t}$
1	**Aim**	Explain a phenomenon P *Answering 'Why P?'*	Realise a practical aim A *Solving a problem, meeting a need*
2	**How to achieve the aim**	Use a scientific theory T *Derive P using T*	Use a technological artefact t *Generate P using X*
3	**Type of understanding**	Scientific understanding of P	Technological understanding of A
4	**How to achieve understanding**	Explain P using T *Construct an interpretation of P by using T*	Realise A using t *Achieve a direct aim a by using t, that embodies the ultimate aim A*
4a	**Step to understanding**	Select or develop T	Select or design t
4b	**Step to understanding**	–	Select P and find X capable of generating P

4. Conclusion

In this chapter, I developed an account of understanding in relation to technology. I proposed the concept of *technological understanding*: the kind of understanding that is involved in the design and use of technological artefacts. I obtained this notion by extending De Regt's (2017) notion of scientific understanding into the domain of technology, and employing an analogy between scientific theories and technological artefacts as a guide. Just as science seeks to understand and explain natural phenomena through the formulation and use of theories, technology seeks to solve practical problems through the design and use of artefacts.

Drawing on this analogy, I argued that the design and use of a technological artefact involve a type of understanding: it requires the cognitive skill to recognise how an

artefact's operation can achieve a particular aim. Just as scientific understanding involves using a theory to explain a phenomenon, technological understanding involves using a technological artefact to realise a practical aim.

Technological understanding parallels scientific understanding, since both involve the cognitive ability to use knowledge effectively – either to achieve explanations (in science) or practical outcomes (in technology). In both cases, such understanding relies on the intelligibility of the respective tools – whether theories or artefacts – to the subject using them.

I further specified the general notion of technological understanding for a design context, arguing that, in this context, technological understanding entails the ability to realise a practical aim by designing an appropriate technological artefact. Specifically, it involves the ability to use knowledge to produce a phenomenon through a physical structure that achieves the aim. Thus, technological understanding is a central epistemic aim in the practice of design.

This account of technological understanding offers a systematic perspective on what understanding means in relation to technology. By arguing that successful engagement with technology involves a cognitive skill – the ability to use a technological artefact to achieve an aim — it brings the epistemic dimension of our engagement with technology into view. Moreover, explicating what this skill amounts to provides a basis for efforts to improve and assess technological understanding. Such efforts become more fine-grained when the contextual character of technological understanding is taken into account.

This chapter specified what technological understanding amounts to in one foundational context: the context of design. The next chapter develops the account further for two additional contexts – practical use and innovation – showing how different contexts of engagement with technology foreground different aspects of technological understanding.

Chapter 3

A Contextual Approach to Technological Understanding

An earlier version of this chapter is under review as: de Jong, E., & De Haro, S. (2026). A contextual approach to technological understanding and its assessment. In: Minds & Machines.

Abstract

Technological understanding is not a singular, fixed concept but a general notion that is specified differently in different contexts. Building on the notion of technological understanding that was developed in the previous chapter – i.e. the ability to recognise how to realise an aim by using a technological artefact – this chapter further develops its pragmatic and contextual character by explicating it as an ability that varies by both context and degree. I complement its original specification for a design context by discussing two additional contexts: operation and innovation. Each context represents a distinct way of realising an aim through technology, resulting in three types (specifications) of technological understanding. To further clarify the nature of technological understanding, I propose an assessment framework based on counterfactual reasoning. Each type of understanding is associated with the ability to answer a specific set of what-if questions, concerning changes to an artefact's structure, performance, or appropriateness – corresponding respectively to a design-type, operation-type, and innovation-type of understanding. Explicitly distinguishing these different types helps to focus efforts to improve technological understanding, clarifies the epistemic requirements for different forms of engagement with technology, and promotes a pluralistic perspective on expertise.

1. Introduction

Having established what understanding essentially means in relation to technology, the previous chapter also pointed to something that demands further elaboration: how context shapes what exactly that understanding amounts to. A taxi driver may 'understand' their car well enough to operate it, but this understanding may prove inadequate when the car breaks down. Similarly, a shipyard owner may have a deep understanding of their vessels' capabilities but lack the understanding required to sail them. Understanding in relation to technology is thus not a single, fixed concept; rather, it varies depending on one's role and goal in relation to a given technology. This chapter elaborates on that contextual character.

Recognising the situated and context-dependent nature of 'technological understanding' has direct implications for efforts to improve people's understanding of a technology, whether it is in education, policymaking or broader public engagement with technological developments. For instance, while the emergence of second-generation quantum technologies has prompted calls to promote public understanding (Vermaas, 2017; Coenen & Grunwald, 2017, p. 292; Seskir et al., 2023; Roberson, 2023; Rathenau Instituut, 2023, p. 3; Nationale Agenda Quantum Technologie, 2019, p. 40), what such understanding entails remains unclear. Does it refer to a simplified, 'light' version of expert understanding, or does it require a qualitatively different kind of understanding? What does it even mean to be an expert in relation to technology?

To address these questions, I take the general account of technological understanding developed in the previous chapter as a starting point and elaborate it for different contexts of engagement. I defined technological understanding as the ability to (recognise how to) realise an aim by using a technological artefact. Emphasising 'use' – which can take shape both as a mental and a practical activity – technological understanding is a *cognitive skill* required for and demonstrated by successfully employing a technological artefact to achieve an aim.

What constitutes 'using a technological artefact' – and, consequently, when technological understanding is achieved – is, within the conditions set by the general notion of technological understanding, relative to context. Different contexts impose different requirements on what counts as sufficient understanding, depending on the agent's role and goals: recall the distinction between the taxi driver and a mechanic or the shipyard owner and a captain. To advance the notion of technological understanding and assess when it has been (sufficiently) achieved, the general concept can and should be further specified for particular, typical contexts.

In chapter two, I took a first step by specifying understanding within the context of technological design. In a design context, the ability to successfully use a technological artefact entails the ability to design it – thinking up and constructing an artefact capable of achieving the desired aim. In this chapter, I further develop the notion of

technological understanding by specifying it for two additional contexts, each representing a typical mode of engagement with technology: 'operation' (practical use) and 'innovation' (devising applications). These three specifications represent distinct, context-relative *types* of technological understanding. Together, they provide a comprehensive account of technological understanding, covering the main contexts in which it applies.[1]

The chapter is organised as follows. Section 2 summarises the account of technological understanding, and further develops some aspects of it: it highlights its pragmatic and gradual character, and distinguishes three main contexts in which technological understanding applies. Section 3 specifies technological understanding for these contexts, resulting in three context-specific types of understanding. To further clarify the nature of technological understanding and its specifications, Section 4 proposes an approach for assessing it based on the ability for counterfactual reasoning.

2. The context-dependence of technological understanding

This section introduces the general concept of technological understanding (Section 2.1) and highlights its contextual, pragmatic and gradual character (Section 2.2). I then expand on this conceptual and gradual character by describing three main contexts in which technological understanding applies (Section 2.3).

2.1 Technological understanding

As argued in the previous chapter, the ability to realise an aim by using a technological artefact involves the cognitive skill of understanding. I drew an analogy with scientific understanding: just like scientific understanding is informally characterised as the ability *to use a theory to explain a phenomenon*, technological understanding is informally characterised as the ability *to use a technological artefact to exploit a phenomenon for a practical*[2] *aim*. In both cases, understanding entails the *cognitive skill* that is involved in using a tool to achieve something – whether giving an explanation or realising practical end.

As noted in the introduction, I understand the term 'use' in a broad sense: it can refer to mental activities, such as employing a theory to explain a phenomenon or recognising how an artefact's capabilities could serve an aim, but also to practical activities, like physically employing an artefact to achieve an aim. 'Using a technological artefact' can thus mean using it conceptually as well as practically. I further illustrate this when I discuss different types of technological understanding.

[1] Subsequent related work appeared in Westerblad (2026) and Zhu (2025).

[2] I call the aim 'practical' since its realisation requires action.

In Chapter 2, I proposed the following *Criterion for Technological Understanding* (CTU):

> **CTU:** An aim A is technologically understood if it can be realised by using a technological artefact t [3]

Building on the accounts of Kroes (2002) and Houkes and Vermaas (2010), I emphasised the dual nature of technological artefacts, as objects with *physical* and *intentional* aspects. The artefact's physical aspect was specified as the combination of a physical structure and a physical phenomenon, and its intentional aspect in terms of the artefact's direct and ultimate aims. Taking the example of a laser, the operation of t (i.e. the laser) fulfils a direct aim (i.e. producing a concentrated beam of light), which in turn serves an ultimate aim (e.g. the ability to cut and weld materials).

Since technological artefacts such as lasers are composite objects – each component of which might also be regarded as an artefact, such as the lens – this raises the question of what should count as *the* artefact in this analysis. I acknowledge the complexity and modularity of many technologies (e.g. cars), which comprise several constitutive 'sub-artefacts' (e.g. a combustion engine, chassis and wheels, automated driving systems). Because technological understanding centres on the ability to *use* an artefact, I define artefacts at the level of use. Therefore, what counts as the artefact depends on the aim: it is the complete device that is employed to realise that aim. Its components form part of its 'inner workings'; though they may themselves qualify as artefacts, relative to a differently framed aim.

Given the formal definition of a technological artefacts given above, explaining technological understanding as the ability to realise an aim A by using t thus means: Realising an aim A by producing a phenomenon P through a physical object X that achieves a direct aim a.[4] This detailed formulation becomes particularly important when I discuss the way in which the artefact must be intelligible in order to speak of technological understanding.

As discussed previously, the ability to use t, which is central to technological understanding, requires that the artefact is *intelligible* to the subject S. This intelligibility was explained as an artefact's 'useability': the combination of features of

[3] This formal criterion highlights that the object of technological understanding – the thing understood – is the aim, and not the artefact itself, just like in scientific understanding the thing to be understood is the phenomenon and not the theory itself. However, as also noted in Chapter 2, since technologically understanding an aim requires qualitative insight into the consequences of the artefact's operation, it can also informally be explained as 'understanding a technology'.

[4] This conception of a technological artefact, and by extension, this account of technological understanding, directly applies to physical artefacts. The account can be extended to apply to software technologies as well, not least because these build on hardware technologies. Expanding the account would, however, require further theoretical elaboration, which is beyond the scope of this thesis.

the artefact that enable the user to recognise its quality as an effective means to achieve an aim A. The user is able to recognise this quality – that is, the artefact is intelligible to them – if they can (counterfactually) reason about the results of t's operation and recognise how to use it to realise the aim A. This counterfactual aspect of intelligibility will be particularly useful in Section 4, when I propose an approach to assessing technological understanding.

To test whether an artefact is intelligible, I proposed the following *Criterion for Intelligible Technological Artefacts* (CITA):

> **CITA:** A technological artefact t is intelligible for a subject S (in context C) if they can recognise qualitatively characteristic consequences of t's operation without practically performing its operation

Both CTU and CITA bring out the pragmatic character of technological understanding: CTU by defining it in terms of the ability to *use* an artefact; CITA by explaining this in terms of the ability to recognise the effects of the artefact's operation. While CTU primarily foregrounds the *object* of technological understanding, CITA makes explicit that understanding always has a *subject* and takes place within a particular *context*.

By bringing out the subject, CITA highlights intelligibility as a relational property, a property of the relation between the subject of understanding and their tool. This subject can be an individual (microlevel), a specific group (meso), or an entire community (macro). Following De Regt's account of scientific understanding, I argue that the meso-level – such as a team – is often the most appropriate level for attributing technological understanding. The account of technological understanding is thus fully compatible with the collaborative character of, for example, design practices or the operation of complex systems.[5] This communal aspect also makes technological understanding an objectively assessable skill. I return to this point in Section 2.2.

Furthermore, by adding 'context', CITA shows that intelligibility is assessed relative to a particular mode of engagement, and that the consequences of an artefact's operation that matter are therefore context-dependent. Because intelligibility is relative to context, technological understanding itself is also context-dependent or 'contextual', that means: what it amounts to and when it is achieved depends on the considered context.[6] I elaborate on both the pragmatic and contextual dimensions of technological understanding in the next sections.

---

[5] The possibility of attributing technological understanding at a meso- or macrolevel is not the same as treating it as a genuinely collective or distributed property. Whether technological understanding can in fact take such a collective form is an open question, one that lies beyond the scope of the present thesis.

[6] I use the terms 'context-dependent', 'context-relative', and 'contextual' interchangeably.

With the formal criteria for understanding in mind, I will from now on primarily use the informal definition of technological understanding – the ability to use a technological artefact to realise an aim, which presupposes the artefact's intelligibility to the user – since this formulation most clearly foregrounds its character as a skill-based capability.

2.2 The pragmatic character of technological understanding

Characterising understanding as a cognitive skill highlights its *pragmatic* nature. It is the ability to *use* knowledge (Reutlinger, Hangleiter & Hartmann, 2018; De Regt, 2017; Grimm, 2012; Wilkenfeld, 2013; Hills, 2016), and thus a cognitive achievement: a successful activity. Similarly, technological understanding emphasises the ability to achieve something: realising an aim by using a technological artefact.

There are three key implications that follow from characterising understanding as a pragmatic notion. First, as a skill, understanding always involves a subject – someone for whom the artefact is intelligible. However, in what follows I will use the term 'agent' rather than 'subject' to emphasise that understanding entails the ability to do something. The implication of an agent highlights that understanding or using knowledge occurs within a specific context: their role and goal determine what constitutes understanding in that situation, making understanding inherently contextual. Moving beyond the individual agent (microlevel), I focus on the 'typical agent' at the mesolevel – an agent representing a class of individuals defined by shared roles and goals, thus constituting a particular context. These roles and goals, in turn, determine the standards for sufficient understanding and the skills required of the typical agent. When technological understanding is achieved is thus ultimately socially shaped. Think here of the requirements for achieving a driver's license. I will further elaborate on the concept of 'context' and the corresponding typical agent – which, in the end, boil down to the same thing – in the next section.

Second, recognising the pragmatic nature of understanding makes it natural to conceive of technological understanding as a *gradual*, rather than all-or-nothing, notion. Understood as a cognitive *skill*, technological understanding can come in *degrees*, as also suggested by other authors (Kelp, 2015; Baumberger, 2019). It is not some kind of knowledge that we have or do not have, but entails the ability to use knowledge, in which we can be higher or lesser skilled. I return to this point in Section 3, where I discuss the three types of technological understanding and indicate what varying degrees of understanding look like in each, and in Section 4, where I explain degrees of understanding more formally in terms of the scope of counterfactual reasoning.

Thinking of technological understanding in terms of graduality aligns with the intuition that it can develop over time, or that one agent may possess a more advanced understanding than another. The idea that understanding can come in degrees can be

traced to the degree of intelligibility: the more intelligible the artefact is, the more (complex) actions the agent can perform using it. Thus, an agent's technological understanding directly relates to the extent to which the artefact is intelligible to them. This gradual character further strengthens the argument for technological understanding as a skill that is, in principle, trainable and testable.

This leads to a third point: technological understanding, conceived as a practical concept, allows for the assessment of its degree based on an agent's ability to perform specific tasks. By conceptualising technological understanding in terms of actions, its observable aspects are highlighted, which provides a foundation for task-based assessment (see Barman et al., 2024 for scientific understanding). This testable nature of technological understanding enables, in principle, the development of metrics and benchmarks. While the aim of this chapter is not to develop a quantitative metric, I will discuss various levels of understanding in Section 3, and explore the possibility of assessing them in Section 4.

2.3 Three contexts of technological understanding

As introduced in Section 2.2, the fact that technological understanding implies a 'user' requires consideration of the context of use, including the goals and skills of the (typical) agent. Thus, technological understanding does not exist in an abstract, context-free form; it is always attained within a particular context and by a particular agent.

By 'context', I refer to a specific mode of engaging with technology, encompassing practices established by a (technological) community. Rather than denoting a particular situation, a context represents a broad category of activities that require the same type of technological understanding. As noted earlier at the end of Section 2.1, whether an agent has attained (a sufficient level of) technological understanding can only be determined in reference to a specific context. This is why I qualify technological understanding as context-dependent or 'contextual'. In this section, I distinguish three key contexts where technological understanding applies.

In addition to the context of design, I introduce the contexts of 'operation' and 'innovation.' These three contexts are distinguished based on three logical possibilities regarding the scope of the general (overarching) concept of technological understanding (Chapter 2, Section 3.2). Beyond the agent, the central elements in technological understanding are the ultimate aim A and the artefact t. The general notion leaves open the question of whether technological understanding includes the selection of A and/or t, i.e. whether these elements are predetermined or fixed. In Chapter 2, I outlined three logical options for the scope of technological understanding. Now I argue that these three logical options represent three typical contexts of engaging with

technology (Table 1) – and that each entails a specific instance or *type* of technological understanding.

In the *maximal account*, neither the aim nor the artefact is given: both are *open*. Technological understanding, in this case, includes the process of selecting a possible aim A and conceiving a suitable artefact t in response. I suggest that this variant of technological understanding is most closely associated with the context of 'technological innovation,' where new connections are created between a specific aim and a technological artefact. Activities aimed at developing innovative technological solutions to a given aim fall within this context, encompassing a wide range of pursuits, from entrepreneurial innovation to innovation policy.

In the *minimal account,* both the aim and the artefact are predetermined: technological understanding does not involve the selection of either the aim A or the artefact t. This does not mean that 'everything is open' – which would pertain to the maximal account – but the opposite: that both the aim and the artefact are 'fixed'. That is: there exists an established relation between a certain aim and an artefact that can (and typically is) used to realise it. This variant typically applies to the context of practical use, where a technological artefact is employed for a common purpose. We encounter this context of practical use whenever we use a technological artefact to accomplish a specific task. I will refer to this context as 'operation.'

In the *via media*, the ultimate aim A is given, but the artefact t is not.[7] In this situation – where the goal is fixed but the artefact is yet to be determined – the focus is on selecting a suitable artefact t to achieve the given aim A. This case is typical of the design context. Like in the previous chapter, I define the context of design as encompassing all practices involved in translating a practical aim into a functional object, as well as the process of physically constructing the artefact (Chapter 2, Section 3.3).

The three logical options for the scope of technological understanding and the three typical contexts to which they correspond are summarised in Table 1 (below).

[7] Technically speaking, there is a second variant of the via media, where the artefact t is fixed, but the aim A remains open. As argued in Chapter 2, I categorise this scenario under the maximal account, as the artefact cannot truly be considered 'fixed' if the aim is unspecified. Formulating an aim entails that the artefact is open for selection – whether by choosing an existing artefact, adapting it to the given aim, or designing a new one. In all these cases, the artefact to be used is not genuinely fixed. If the aim A remains open, the artefact t is likewise open. Conversely, selecting an artefact does not necessarily entail the formulation of an ultimate aim.

Table 1. Three typical contexts in which technological understanding applies, depending on whether the aim and the artefact are open (underdetermined) or fixed (predetermined)

Logical option		*Aim*	*Artefact*	*Context*
(i)	maximal account	open	open	**innovation**
(ii)	minimal account	fixed	fixed	**operation**
(iii)	via media	fixed	open	**design**

Each of these three contexts involves a type of technological understanding, in which the ability to use a technological artefact is specified in a distinct way. These types are developed in more detail in Section 3. For now, it is important to see that technological understanding is always realised within a specific context, and that there is, in this sense, no 'general' technological understanding over and above its context-specific manifestations. The general notion of technological understanding serves to set the conditions for what counts as technological understanding.

2.3.1 Intelligibility and its relation to degrees of understanding

Intelligibility is a prerequisite in all contexts for the successful use of the artefact t. Recall that for an agent to use the artefact t, t must first be intelligible to them. Intelligibility is thus a qualification of the relation between a subject and an artefact. The criteria for this qualification vary with context: whether an artefact is deemed to be (sufficiently) intelligible depends on the specific context. Different contexts and aims impose varying demands on the intelligibility of a technological artefact.

This can be most clearly understood in terms of the components of the artefact t that are emphasised by the notion of technological understanding in a given context – specifically, the physical structure X, the phenomenon P, the direct aim a and the ultimate aim A. These components interact differently depending on the context of use, which means that the artefact t may be (primarily) intelligible in terms of X and P, or in terms of a or A.

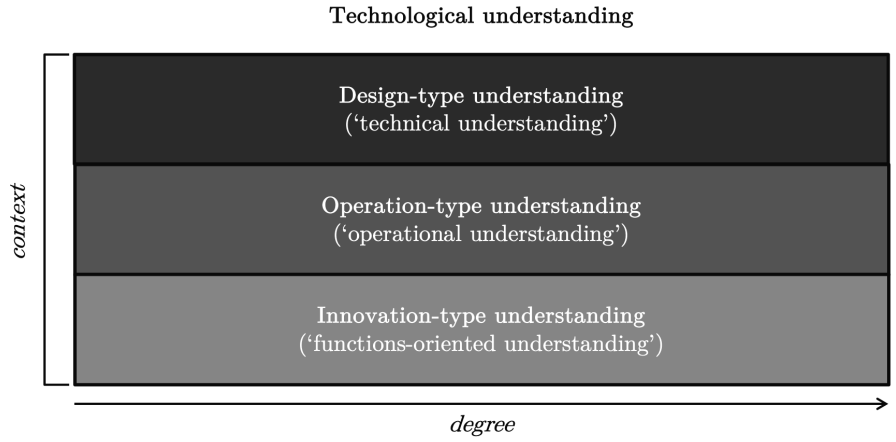
This distinction pertains to the focus of counterfactual reasoning involved (see Section 4). Technically, the direct aim a emerges from the interplay between X and P; however, by focusing on a, the phenomenon is framed in terms of a specific objective. The focus is shifted from, for example, the technical details of a laser's operation (i.e. stimulated emission) to its practical result (i.e. a concentrated light beam). If, given the context, the direct aim a is the primary consideration for the intelligibility of t, then the agent will predominantly derive the consequences of the artefact's operation in terms of a, with less emphasis on its physical structure X and the phenomenon P it produces. In

contrast, another context may require the artefact's intelligibility at the level of both X and P, in which case the agent will reason about how the interaction between X and P can be used to achieve the ultimate aim A.

Within any given context, the intelligibility of a technology can be understood as existing on a continuum. This means that intelligibility comes in varying degrees, which in turn reflects different levels of technological understanding. Each type of technological understanding, therefore, is an ability that can be possessed to varying extents and should be treated as a spectrum.

The threshold for what constitutes 'sufficient' understanding is contingent upon the specific context. For instance, in the context of operation, the level of technological understanding required for professional use may differ significantly from that required in a private setting. Therefore, the criteria for defining technological understanding are inherently context-dependent, with varying thresholds depending on the role and goals of the typical agent. As discussed earlier, specific roles and goals in engaging with technology can also vary within a given context. *Figure 1* illustrates the various contexts and their corresponding types of technological understanding, emphasising that the notion that understanding exists on a spectrum.

Figure 1. Three specifications of technological understanding in three different contexts



2.3.2 Consequences for expertise

The contextual and gradual view of technological understanding – as an ability that varies in both type and degree – has significant implications for the concept of expertise. If expertise is conceived as possessing an advanced level of understanding, then, given the context-dependence of understanding, expertise itself is also context-dependent. The idea that different types of technological understanding apply to

different contexts allows for multiple interpretations of expertise, shaped by the specific context.

Beyond context, the roles and goals of an agent influence what qualifies as expertise. For instance, in an operational context, the same level of understanding might be seen as expert in one scenario but insufficient in another. Just as the threshold for 'sufficient understanding' varies across roles, so does the definition of expertise; what qualifies as 'advanced' understanding is role- and goal-dependent.

This context-dependent, agent-relative view of expertise aligns with common intuitions: a designer may be considered an expert in a laboratory setting but not necessarily in everyday operational contexts. Distinguishing between different types of technological understanding thus invites a reconsideration of what it means to be an expert.

2.3.3 Unrealised aims: A trip to Mars

One interesting question raised by this account of technological understanding is whether aims that have not yet been realised – or may never be realised – can nonetheless be considered technologically understood. As a paradigmatic example, consider a manned trip to Mars: detailed plans and designs exist, yet the mission has not been carried out, and it is uncertain whether humanity will ever achieve it. Can there be technological understanding of such an aim in the absence of actual achievement?

In short: not really. Realisation, or 'success', is a condition for treating an aim as technologically understood. Since technological understanding is defined as a skill, there must be some form of demonstrability. In the case of a trip to Mars, the aim has not yet been realised in practice, so it must be concluded that technological understanding of this aim has not yet been achieved.

It is important, however, to clarify the conditions of success. The demonstration of technological understanding need not always be practical; it may also be conceptual, provided that the functional principles are worked out in a defensible, grounded, and theoretically plausible way. Technological understanding may be attributed as a skill even if it has not yet been exercised in a concrete, practical way, provided that there are sufficiently strong grounds for expecting success under specified circumstances – where such 'sufficiently strong grounds' are to be assessed by the relevant expert community. We may attribute technological understanding to a pilot who recognises how to land in an emergency situation without ever having faced one, as well as to an engineer who is able to build a bridge design that has not yet been constructed.

If the case of a manned mission to Mars lacks *practical* demonstration of technological understanding, can at least *conceptual* realisation of its aim be claimed? Not unequivocally. Unlike the pilot, whose competence lies within a well-practiced and

extensively validated domain, a Mars mission entails performance under largely untested conditions. The step from current designs and simulations to operational success is therefore much larger and more uncertain. In the absence of sufficient practical validation, the grounds for expecting success remain contestable. Conceptual realisation of technological understanding in the Mars case thus cannot yet be securely attributed – though it cannot be denied with confidence either. I revisit this issue in 3.3.2.

Furthermore, the Mars example illustrates that technological understanding is subject-relative and context-dependent. As with scientific understanding, technological understanding can be achieved at different subject-levels: micro (an individual engineer), meso (a team or user group), and macro (society or humanity). Technological understanding may be considered generalisable in the sense that realisation at any of these levels is sufficient for claiming that an aim is technologically understood, even if some agents or contexts have not attained it. Whether an unrealised aim can nonetheless be regarded as technologically understood therefore depends on the level at which realisation is assessed: we may ask whether technological understanding exists in general (because the aim has been realised by one agent in a given context), or whether it has been attained in a particular context or by a specific group.

At the end of Section 3, I revisit the Mars example to explore whether the success-condition applies equally to all three kinds of technological understanding, or whether the aim to visit Mars may instantiate one specific type of understanding.

3. Specifying technological understanding: three types

Drawing on the distinction between the three contexts of technological understanding, this section further elaborates on the three corresponding types of understanding. I first briefly summarise the design-type of understanding (3.1) and then develop the operation-type (3.2) and the innovation-type (3.3). These context-specific types of technological understanding should be understood as context-relative specifications of the general notion, which is concretised in each of these types. In unpacking each type, I show how different forms of intelligibility are required, and how degrees of technological understanding emerge within each context.

3.1 Design-type technological understanding

Technological understanding requires the availability of an artefact that can realise that aim. This is, however, not a given. Often there is only an aim and the desire for an artefact that can achieve it. In this case, 'using an appropriate artefact' requires *designing* one (logical option (iii), *via media*).

More formally, successfully designing an artefact t involves the ability to construct a physical structure X that produces a phenomenon P that achieves a direct aim a, which realises the ultimate aim A. In a design context, technological understanding is thus instantiated as the ability to develop an artefact suited to a given practical aim. Conversely, the ability to design an appropriate artefact both requires and reflects technological understanding.

The type of technological understanding involved in the design context was thus defined as follows:

> **Technological Understanding(design)** is the ability to realise an aim A by *designing* a technological artefact t

3.1.1 Intelligibility at the level of inner workings

Achieving technological understanding requires the artefact to be intelligible. In a design context, this means the ability to reason about the artefact at a technical level. A designer must recognise how an artefact's consequences stem from its design. The design-type of technological understanding may therefore also be colloquially referred to as 'technical understanding'.

From the perspective of technological understanding's key components, technical intelligibility emphasises the ability to derive the artefact t's consequences in terms of its structure X and the active phenomenon P. In other words, if intelligibility concerns knowing how t can be used to achieve A, then in a design context, it means understanding how the specific relationship between X and P contributes to A.

Since intelligibility exists on a spectrum, different levels of design-type understanding can be distinguished. The required depth of intelligibility depends on an agent's role and goals. For example, for an electrical engineer designing a technological artefact (e.g. a smartphone or a lift), the artefact needs to be intelligible to a more advanced extent than for a technician responsible for its implementation or repair.[8]

3.2 Operation-type technological understanding

The second context is that of practical use, or 'operation', corresponding to the option where both the aim and the technological artefact are relatively fixed (option (ii), minimal account). By 'practical use,' I refer to the direct operation of a technological artefact with the intention of achieving a specific practical aim. Here, 'using an artefact'

[8] It is true that practical implementation of technology requires additional knowledge and skills pertaining to specific contextual factors, but the required depth of knowledge about the physical structure and the phenomena that compose a given artefact is less than for design practices.

is specified as operating it. This leads to a second specification of technological understanding, this time within the operational context:

> **Technological understanding(operation)** is the ability to realise an aim A by *operating* a technological artefact t

It is important to highlight from the outset that the operational context is highly diverse. It encompasses everything from the professional use of complex machines – such as MRI scanners, industrial lasers, or cranes – to the everyday operation of simpler devices by consumers, such as electric toothbrushes, computers, or cars. However, as discussed in Section 2.3, a unifying perspective emerges when considering the type of technological understanding involved in these cases. Despite the variety of technologies and their uses, all forms of operation rely on a specific kind of technological understanding. While such differences may be relevant, for example, in the context of setting up licensing for use, this chapter will not distinguish between different types of operation, though I acknowledge that such distinctions exist. Instead, I will primarily consider these differences in terms of varying degrees of the operation-type of understanding, or in short 'operational understanding'.

I should also emphasise that operating an artefact typically requires additional practical skills, tailored to the specific circumstances that affect the artefact's operation. Technological understanding encompasses the epistemic dimension of the activity: the cognitive ability to recognise how an artefact can be operated, rather than the physical ability to do so. For example, operating a laser involves understanding how the laser can be used to cut material. While this cognitive understanding is necessary, it is not always sufficient for successful operation, as physical handling skills are also required. Therefore, successful operation demonstrates both cognitive understanding and physical skills, with the latter not being part of the cognitive skill itself.

3.2.1 Intelligibility at the level of direct results

In the operation context, an artefact t is intelligible for a user if they can reason about it at an *operational* level – enabling them to practically use the artefact. For an artefact to be operationally intelligible, an agent must primarily recognise its consequences in terms of how it achieves the direct aim a, and ultimately, the broader aim A. The artefact only needs to be 'technically' intelligible (in terms of X and P) to the extent that this understanding is necessary for its successful operation.

This level of practical reasoning differs from the technical reasoning required by designers. For practical use, extensive knowledge of the physical phenomena underlying an artefact's operation is often not necessary. For instance, one can send an email without understanding how a CPU or the internet works. However, in some cases, reasoning about an artefact and its operational consequences does require some insight

into its internal mechanisms. Operating a crane requires an understanding of gravitational forces. The required depth or quality[9] of this insight depends on the specific role and goals of the agent. Ultimately, the user needs to understand enough about the inner workings of the artefact to recognise how it can be used successfully.[10]

Operational intelligibility should also be understood as a matter of degree: for an advanced user, the artefact will be more intelligible than for an inexperienced one, resulting in a higher or lower level of operational understanding. The required level of understanding depends on the agent's role, goals, and the specific circumstances. For example, driving a car on flat tarmac may require less technological understanding than driving off-road. Similarly, using a computer in a professional context generally demands a higher level of proficiency than using it for personal tasks.

3.3 Innovation-type technological understanding

The third context that was distinguished is the context of innovation, where both the ultimate aim A and the artefact t are open (option (i), maximal account). While the designer's role is to identify an artefact that can achieve a given aim, and the end-user operates the artefact to achieve a specific aim, the innovator defines the aim to be achieved and conceptualises the technological response by selecting (or adapting) an existing artefact or conceiving a new one yet to be designed.

This requires a deeper, 'embedded view' of the ultimate aim than is needed in design or operation contexts. The innovator must possess a comprehensive understanding of the aim's nature and its societal implications, enabling them to assess whether, and to what extent, a technological artefact can fulfil it.

While the concept of innovation lacks a universally accepted definition (Bontems, 2014), it is generally associated with novelty and challenging the status quo. The Cambridge Dictionary defines innovation as "the creating and use of new ideas or methods," while the Oxford Dictionary describes it as "the process through which new products, concepts, services, methods, or techniques are developed." Furthermore, innovation is frequently linked to the creation and application of technology (Bontems, 2014; Blok & Lemmens, 2015; OECD, 2005).

---

[9] Is it, for example, enough to know that something is the case (e.g. knowing that gasoline will destruct my diesel engine) or is it required to understand why something occurs? For technological understanding, knowing-that is not enough: one should be able to see why something is the case. However, the range of why-questions that one should be able to answer, depends on the circumstances. I expand on this idea of technological understanding as the ability to answer why-questions in Section 4.

[10] I adopt the view that, generally, technological understanding requires scientific understanding (for a more detailed discussion, see Chapter 2). Both notions come in degrees and have contextual standards, so the depth of the required scientific understanding is attuned to the type and degree of technological understanding.

A comprehensive discussion of innovation is beyond the scope of this thesis (for a philosophical exploration of the concept, see, for example, Blok, 2021). Instead, I adopt a broad conception of innovation as the ideation of a technological response to a desire to change the status quo. This definition views innovation as an activity rather than a specific invention, focusing primarily on the ideation of opportunities for the successful application of a technological artefact rather than its actual creation.

This conception of innovation incorporates elements of both the design context and the context of practical use. Like a designer, an innovator establishes a connection between an aim and an artefact that did not previously exist. However, unlike the designer, the innovator does not directly create or design the artefact. Instead, the innovator 'designs' the relationship between the artefact and the aim. This involves conceptualising, outlining, or describing in broad strokes a potential relationship. In essence, the innovator does not design the technological artefact itself but focuses on devising how a (yet to be designed) artefact might be applied to achieve a specific aim.

Like the end-user, the innovator is primarily focused on the ultimate aim A and considers the artefact t as a potential means to that end. Successfully doing so requires not only insight into the aim but also an understanding of the sociotechnical context in which the artefact must function (de Jong, 2022: pp. 14-16). However, unlike the end-user, the innovator is not necessarily the one who will practically use the artefact and does not follow a pre-existing t-for-A relationship. In fact, the innovator determines the problem, need, or demand and conceives a technological response to address it.

The innovator 'devises' the artefact as a potential means to an aim, with 'devising' referring to the process of envisioning or ideating a technological response (e.g. the use of MRI) to address a specific need (e.g. non-invasive examination of soft body tissues). This means they must be able to align the artefact's capabilities with an aim and assess its performance in relation to that aim. The innovation-type of technological understanding is therefore capability-oriented, focusing on the functionalities and potential applications of the artefact. This type of understanding may therefore also be referred to as 'functional understanding'[11].

Innovators may conceive entirely new technologies, such as generative Artificial Intelligence and quantum computers, or discover novel applications for existing artefacts. For instance, radar communication systems were repurposed to heat food, leading to the invention of the microwave. Similarly, Magnetic Resonance Imaging

[11] 'Functional' here denotes an understanding of a technology's functional capabilities, that is: its potential functions and applications. This should not be taken to imply that other types of understanding lack functional character; understanding is inherently a pragmatic notion and is, in that sense, functional by definition. 'Functional understanding' should thus be read as shorthand for 'functions-oriented understanding', and I prefer the former here for readability.

(MRI), originally designed for medical diagnostics, has been adapted for food quality assessment. Also, before the combustion engine was invented, innovators attempted to repurpose railway steam engines for early steam-powered cars. Innovators play a key role in (re)purposing and (re)imagining technological artefacts, extending them into new domains and applications.

Considering the innovator as a typical agent provides us with a third specification of technological understanding, specifically applicable to an innovation context:

> **Technological understanding(innovation)** is the ability to realise an aim A by *devising* a technological artefact t

3.3.1 Intelligibility at the level of functional capabilities

For an innovator to devise an artefact and connect its capabilities to a specific aim, the artefact must be intelligible at an appropriate level. In an innovation context, intelligibility entails the ability to reason about the artefact in an evaluative manner: like the operator, the innovator must recognise the artefact's consequences, particularly in terms of how effectively the achievement of a direct aim a contributes to an ultimate aim A. However, the innovator's required insight into the problem, need, or demand that the artefact is meant to address is necessarily deeper.

At this level, the artefact must be intelligible in terms of its functional properties rather than its inner workings or immediate, concrete effects. To invent a use case, the innovator must grasp the artefact's capabilities and potential applications. This aligns with Houkes and Vermaas's (2010) function analysis, which defines ascribing a function to an artefact as assigning to it a specific (physicochemical) capacity (p. 84). Functional intelligibility can therefore be understood as *epistemic access to the artefact's capabilities*[12]. In some cases, depending on the agent's role and goal, devising a technological artefact may also require insight into how these capabilities emerge from its design and use.

The degree of functional intelligibility required varies according to the agent's role, goals, and circumstances. A strategic visionary – who conceives a novel relationship between an artefact and an aim – may require a different depth of understanding than a normative assessor, who evaluates the artefact's use, not only in relation to its intended aim but also within a broader societal or ethical framework. Each of these roles demands a tailored level of intelligibility and, consequently, a corresponding degree of understanding.

[12] I favour the use of the term 'capabilities' rather than 'capacities', since 'capabilities' refers to the qualitative aspects of a technology's potential abilities and functionalities, whereas 'capacities' is often associated with the quantitative limits or extents of these abilities.

Now that I have outlined three types of technological understanding, I should emphasise that they are not competing or mutually exclusive; instead, they are complementary and in practice, overlap or intertwine. A user may demonstrate innovation by employing an artefact in unforeseen ways, an innovator may engage in design by co-formulating design requirements, and a designer may enter the sphere of innovation by (co)determining the ultimate aim. This complementarity also explains why an agent might possess high understanding of one type while having low or no understanding of another: the different types of technological understanding track different dimensions of our engagement with artefacts, that can perfectly coexist. The conceptual distinction between these types primarily serves to highlight which form of understanding primarily applies in a given context.

Beyond clarifying these distinctions, this framework also has normative significance. It provides a basis for arguing that certain types of technological understanding are necessary to undertake specific actions within each context. Moreover, within any given context, the degree of technological understanding required depends on the agent's role and goals. This perspective – that different types and degrees of understanding are required for different actions – underscores the need for a method to assess technological understanding. In section 4, I propose one possible approach for this.

3.3.2 Revisiting the Mars example in light of the three contexts

Recall the example discussed at the end of Section 2: the unrealised aim of a manned trip to Mars. I concluded that, by definition, an unrealised aim cannot be considered to be technologically understood, since realisation is a necessary condition for understanding: technological understanding is achieved in the realisation of an aim.

One might, however, object that in the Mars example there is indeed no design-type understanding (yet), and, due to the absence of a well-functioning artefact, also no operation-type understanding – but that this does not necessarily preclude innovation-type understanding. Depending on the quality of the plans, the Mars case could demonstrate an innovation-type of understanding – that is, understanding the prospective functionalities of a future artefact in the light of a given aim, even if design- and operation-type understanding remain absent. Claiming innovation-type understanding of the aim to set foot on Mars would then exemplify a *conceptual* realisation of technological understanding.

However, conceptual realisation is not equivalent to mere conceivability; it is conditional on an epistemic warrant (see 2.3.3). Claiming an innovation type of understanding in the absence of a functioning artefact is difficult, because it cannot be epistemically confirmed. Until the functionalities are specified and warranted in an operationally grounded manner, we remain in the realm of speculation rather than understanding. In the absence of such evidence, as in the case of a manned trip to

Mars, we cannot yet claim an innovation-type of understanding. Such a claim can only be made once a manned Mars mission (or various missions of comparable difficulty) becomes a reality, it becomes possible to retrospectively justify such a claim and assess previous plans as indeed demonstrating innovation-type understanding.

4. Assessing technological understanding

To clarify the nature of the different types of technological understanding, I propose an approach for assessing it through counterfactual reasoning (4.1), building on a similar benchmark for scientific understanding (Barman et al., 2024), and argue that different types of technological understanding correspond to distinct sets of relevant counterfactuals (4.2). This discussion directly builds on the contextual character of technological understanding. If what technological understanding amounts to depends on the context – within the bounds set by the general notion – then it can be meaningfully assessed only relative to a specific context, since the context determines which counterfactuals matter. Rather than weakening the concept, this context-relativity offers precisely the anchor points needed for empirical assessment.

4.1 Technological understanding as a form of inferential reasoning

The pragmatic nature of understanding – as the cognitive skill to use knowledge – suggests that it can be assessed through an agent's ability to perform specific tasks. Technological understanding is evidently demonstrated by the successful use of a technological artefact. However, to further illuminate the cognitive dimension of this achievement, I here consider the cognitive tasks involved in this process. To do so, I once again draw an analogy with scientific understanding.

The argument for scientific understanding is as follows:

1a) *Proper use of a theory as the core ability:* In the case of scientific understanding, the fundamental cognitive task is knowing when and how to apply a scientific theory to explain a phenomenon. More specifically, scientific understanding involves recognising the conditions under which a given theory is relevant and how it can be used to generate explanations. For example, understanding the motion of an object under different forces means knowing when and how to apply Newton's laws to predict its motion under various conditions.

1b) *Proper theory use requires inferential reasoning:* The ability to properly use a scientific explanation depends on the capacity to draw counterfactual inferences (Woodward, 2003; Woodward & Hitchcock, 2003; Weslake, 2010; Grimm, 2012). Building on this, Barman et al. (2024) argue that scientific

understanding[13] can be assessed through an agent's 'inferential ability' – that is, the ability to infer how the phenomenon P would behave according to a theory T given specific circumstances. Thus, scientific understanding, understood as the ability to use a scientific explanation, presupposes and can therefore be tested by an agent's ability to reason counterfactually.[14]

1c) *Inferential reasoning can be tested with what-if questions:* Barman et al. (2024) argue that the inferential ability central to scientific understanding can be evaluated by an agent's capacity to answer a specific type of question – namely, questions about what would have happened if certain factors were different or what would happen if specific changes were made. Successfully answering such 'what-if-things-had-been-different' and 'what-would-happen-if' questions is a key indicator of scientific understanding.[15,16] This ability involves "postulating hypothetical scenarios about what would occur under a specific set of circumstances" (Barman et al. 2024, p. 5).

Their account builds on Kuorikoski and Ylikoski (2015), who argue that "correct what-if inferences provide a natural measure of understanding" and suggest that the breadth and precision of an agent's counterfactual inferences correlate with their degree of understanding: the greater the range and accuracy of such inferences, the deeper the level of understanding.

By analogy, I propose the following argument for the cognitive skill constituting technological understanding:

[13] Based on the account by De Regt (2017).

[14] While Barman et al. (2024) primarily focus on correctly applying a theory under different conditions, counterfactual reasoning also plays a crucial role in evaluating how modifications to the theory would impact its explanatory or predictive power (Woodward, 2003). In other words, counterfactual reasoning is not only essential for applying scientific theories but also for their refinement and adjustment.

[15] More precisely, Barman et al. argue that the degree of an agent's scientific understanding of a phenomenon P can be determined by assessing the extent to which the agent "(i) has a sufficient complete representation of P; (ii) can generate internally consistent and empirically adequate explanations of P; (iii) can establish a broad range of relevant, correct counterfactual inferences regarding P" (2024, p. 4). Together these abilities constitute scientific understanding; they do not represent subsequent levels of understanding. Barman et al. propose to measure these abilities through answering what, why, and what-if-things-had-been-different questions. Since the ability to answer what-if questions builds on – and therefore, includes – the ability to answer what- and why-questions, I will focus on the former as the primary indicator for understanding. This aligns with what Barman et al. refer to when they state that "the range of counterfactual inferences an agent can articulate is strongly related to the level of understanding". (p. 5)

[16] For convenience, I will further refer to what-if-things-had-been-different questions and the related 'what-would-happen-if-questions' simply as what-if questions.

2a) *Proper*[17] *use of an artefact as the core ability:* Just as scientific understanding is centred on the proper use of a theory, technological understanding fundamentally entails the proper use of an (adequate) technological artefact, i.e. knowing when and how to use it to achieve an intended aim.

2b) *Proper artefact use requires inferential reasoning:* Given the structural parallels between scientific and technological understanding, the cognitive skill involved in using a technological artefact can also be understood as an inferential ability – specifically, the ability to draw inferences about the artefact's application. This means being able to predict how the relationship between the direct aim a and the ultimate aim A if the artefact t were to be used in specific circumstances. This involves reasoning about when and how to apply the artefact under changing circumstances, and how changes to the artefact itself could affect its performance.[18] Both perspectives help to improve artefacts to better achieve their intended aims, analogous to refining a scientific theory to enhance its predictive power. For example, successfully using a microwave to heat food requires, depending on the context, understanding how to use it under different conditions (e.g. adjusting power levels for different types of food) and how modifying the microwave (e.g. redesigning its heating elements) would impact its performance.[19]

2c) *Inferential reasoning can be tested with what-if questions:* The ability to make what-if inferences – both regarding changing circumstances and modifications to an artefact's design – is a key indicator of technological understanding and can be assessed through what-if questions. Just as what-if questions about a phenomenon P test scientific understanding by gauging an agent's ability to explain P under different conditions using a theory T, I argue that technological understanding can be tested based on an agent's ability to realise an ultimate aim A under varying circumstances using an artefact t.

---

[17] I use the term 'proper' in analogy with arguments in the literature on scientific understanding, understanding it as 'successful'. I am aware of the fact that 'proper use' of a technological artefact is a site of debate within the philosophy of technology. Rather than possessing a fixed identity that prescribes a correct mode of use, there is a strand of literature that stresses the interpretative flexibility of technological artefacts, leaving room for users to 'appropriate' them in different ways (Bijker & Pinch, 1987; Carroll, 2004; Verbeek, 2006). In this context, talk of 'proper' or 'improper' use is less accurate; it is more common to distinguish between 'intended' and 'unintended' uses. Here, I am concerned with uses of an artefact that are conducive to achieving a given aim – hence 'successful'. Defining what counts as successful, however, is not reserved to a fixed group, for example those who designed the artefact; success conditions are socially constructed and vary upon context.

[18] In Section 4.2, it is explained that the inferential ability that focuses on changes in the artefact particularly pertains to the design-type of technological understanding.

[19] This typically applies to the context of design and possibly to the context of innovation.

These include questions about how an aim can be achieved using an artefact under different conditions, for example how effectively a microwave (t) heats food (A) when heating a fresh versus a frozen meal, as well as questions about how changing the artefact or its use affects its performance. For instance, how adjusting the heating time, altering power output, or redesigning the microwave's internal structure influences its ability to heat food.

The ability to infer what would happen if certain variables were changed is directly tied to the ability to bring about a desired outcome. This is why what-if questions can often be reformulated as *how* questions: *How can* t *be used to achieve* A *under a specific set of conditions?* This manipulative aspect of inferential reasoning (Woodward, 2003; Kuorikoski & Ylikoski, 2015) is fundamental to technological understanding, which is inherently manipulative, geared towards bringing about a certain state of affairs. In this sense, technological understanding is not merely indicated by correct what-if inferences, but *consists* in them.

The breadth and depth of an agent's technological understanding is indicated the number and complexity[20] of what-if questions they can answer correctly. The greater their ability to establish counterfactual inferences, the higher their degree of technological understanding. However, as Barman et al. (2024) acknowledge, the specific questions deemed relevant for assessing understanding are context dependent. Similarly, determining a *sufficient* level of understanding varies across different situations, depending on the demands of the task at hand.

To sum up, technological understanding, like scientific understanding, involves inferential reasoning and can be assessed by an agent's ability to answer what-if questions about the use and modification of technological artefacts under varying conditions. Since technological understanding is defined by the ability to answer what-if questions, its context-dependent nature implies that different contexts require different sets of relevant questions. I explore this further in the next section.

4.2 Inferential abilities per context

In the previous section, I argued that technological understanding can be assessed through the ability to answer what-if questions, which involves generating scenarios and reasoning about how an aim can be achieved by using an artefact under varying circumstances. Since each type of technological understanding corresponds to distinct abilities – namely: successfully designing, operating, or devising an artefact – different counterfactuals are relevant in each case. In this section, I specify the set of what-if questions pertinent to each type of technological understanding. In doing so, it should

[20] Complexity here refers to the generality of a question, where more general questions are associated with lower levels of complexity and more specific questions with higher levels.

be kept in mind that standards for intelligibility and, by extension, standards for understanding, are ultimately established within technological communities and practices.

4.2.1 Assessing operation-type technological understanding

Successfully operating an artefact requires understanding how to use the artefact under varying conditions, specifically how changes in those conditions impact its performance and the achievement of the intended aim. This operation-relevant inferential ability can be assessed through what-if questions that explore how to achieve an aim under different circumstances by operating the artefact in specific ways. This includes questions like: *What if the operating conditions changed? How would this affect both the performance and the proper operation of the artefact?* and *What would happen if the artefact malfunctioned?* For example: *How do different weather conditions affect my car's performance, and how can I adjust my driving to stop at the traffic sign in time?*

Answering such questions requires operational intelligibility: the artefact needs to be intelligible at the level of its direct practical results. The artefact is intelligible at this level if the operator can recognise its consequences in terms of the direct aim it realises (e.g. pressing the brake pedal results in slowing the car). However, answering operation-relevant what-if questions may also require some degree of technical intelligibility of the artefact. For example, to operate a crane properly, a crane driver needs not only to understand how weather conditions affect the lifting of an object but also how the angle of the slings influences the maximum weight that can be safely lifted.

4.2.2 Assessing design-type technological understanding

Successfully designing an artefact requires understanding how changes to the artefact's structure or components affect its performance. This design-relevant inferential ability can be assessed through what-if questions that explore how an aim can be realised under specific circumstances by designing the artefact in particular ways. In a design context, the relevant what-if questions focus on how the artefact's design can achieve a given aim under varying conditions: *What would happen if the artefact were designed differently (e.g. by using different materials)? How would this affect its performance in relation to the aim?* For example: *What would happen if we replaced the red laser in this machine with a blue laser (shorter wavelength)? How would this affect the machine's measuring accuracy?*

Answering such questions requires *technical intelligibility*: the artefact must be intelligible at the level of its internal structure and functioning. As discussed in Section 3, an artefact is intelligible at this level if the agent can recognise the consequences of its operation based on its physical structure X and the active phenomenon P. Design-

relevant what-if questions thus focus on deriving the artefact's consequences based on specific configurations of X and P.

4.2.3 Assessing innovation-type technological understanding

Successfully devising a technological artefact involves linking the capabilities of technologies to practical aims. This requires understanding how and to what extent an artefact's potential capabilities can realise a specific goal. Within this context, relevant inferences involve reasoning about how new or existing technologies can be applied to solve problems or achieve other objectives. This includes questions like: *What would happen if we changed the aim – how would that affect the appropriateness of the artefact as a technological response?* and *What would happen if we changed the artefact (partly or entirely) – how would that affect the achievement of the aim?* For example: *What if we integrated AI-driven diagnostics into healthcare services? How would the capabilities of AI enhance the effectiveness and reach of these services? Or what if we applied autonomous vehicle technology to public transportation systems in congested cities? How would the features of autonomous vehicles contribute to the aim of reducing traffic and improving commute times?*

Answering such questions requires functional intelligibility: the agent must recognise the artefact's capabilities and assess their relevance to specific goals ('fitness for purpose').

In an innovation context, the relevant set of what-if questions may also address the design and operation of the artefact if such questions help determine whether the artefact is an appropriate technological response to the aim. Like in the operational context, the focus is primarily on the relationship between the aim and the artefact, though in some cases, reasoning about the artefact's consequences at the level of its inner workings may be necessary for successful devising.

Table 2 summarises the core abilities for each of the three types of technological understanding (row 1), the corresponding level of intelligibility (row 2), the counterfactual reasoning indicative of each type (row 3), and paradigmatic what-if questions (row 4).

Table 2. Three types of technological understanding, their associated level of the artefact's intelligibility, and the inferential abilities by which they can be assessed

	Type of understanding	*Operation-type*	*Design-type*	*Innovation-type*
1	Technological understanding is the ability to realise an aim by…	…**operating** a technological artefact	…**designing** a technological artefact	…**devising** a technological artefact
2	Requires the artefact to be intelligible at the level of its…	practical results	inner workings	functional capabilities
3	Assessable by testing the inferential ability to recognise how…	…different circumstances affect the operation of the artefact	…changes in the internal structure of the artefact affect its performance	…appropriate and effective the artefact is with respect to a chosen aim
4	What-if questions that can test the inferential ability	*What if the operating conditions changed?* *How would this affect the performance as well as the proper operation of the artefact?*	*What would happen if the artefact was designed differently?* *How would this affect its performance with respect to the desired aim?*	*What would happen if we changed the aim or the artefact?* *How would that affect the appropriateness of the artefact as a technological response?*

5. Conclusion

This chapter built on the idea that every (successful) engagement with technology involves a form of understanding – a cognitive skill that goes beyond the mere possession of knowledge. It showed that what counts as (sufficient) technological understanding depends on the context: on how we relate to a technological artefact (our role) and what we wish to achieve (our goal). In this chapter, I further developed the notion of technological understanding as an ability that varies by context and degrees. I complemented its specification for a design context – as the ability to design a technological artefact – with two additional contexts in which technological understanding is involved: operation and innovation. In these contexts, technological understanding is specified, respectively, as the ability to practically use a technological artefact and to devise its applications. Together, these three contexts capture the main modes of engaging with technology that involve technological understanding. This

yields an account that distinguishes three context-relative manifestations, or types, of technological understanding – all situated within, and constrained by, the conditions set by the general notion.

Each type comes in degrees and each requires that the artefact is intelligible to the agent, albeit at different levels. The design-type requires an understanding of the artefact's inner workings, the operation-type focuses on its practical results, and the innovation-type centres on its functional capabilities.

Further unpacking the pragmatic character of technological understanding as an ability, I have argued that it consists in – and can be assessed through – the competence to establish counterfactual inferences. This counterfactual competence lies at the core of all three types of technological understanding; what changes across contexts is which counterfactuals count as relevant.

I suggested that design-type understanding can be assessed through what-if questions about changes in an artefact's internal structure, operation-type through what-if questions about how varying circumstances affect its performance, and innovation-type through questions about an artefact's effectiveness and appropriateness for a specific aim. This framework provides not only a method for assessing technological understanding but also a foundation for articulating required degrees of understanding and evaluation criteria in specific contexts; standards that are ultimately socially established within technological communities and practices. The next chapter builds on this idea as it explores the kind of technological understanding that grounds discussions about societal impact of technology.

Although the distinctions between the design-, operation-, and innovation-type of understanding may seem intuitive, it is important to explicitly distinguish them, since they are easily conflated in practice. Such conflations can misguide efforts aimed at improving technological understanding and wrongly (dis)qualify certain forms of understanding. Conceptually distinguishing them – while acknowledging their possible overlap – sharpens our focus on the required type of understanding, for example when discussing a technology's societal impact (see Chapter 4). Moreover, explicitly recognising different types of technological understanding fosters a more pluralistic perspective on expertise in relation to technology, in which multiple types of expertise are acknowledged and valued. Thus, technological understanding is not a one-size-fits-all skill; it is a contextual competence and essential for any successful form of engagement with technology.

As a novel concept, technological understanding naturally raises a range of further questions, not least because 'understanding' itself is a much-debated notion. These include, for example, whether technological understanding can occur at a collective level, what it implies for (the distinction between) proper and improper use of technological artefacts, and how the different types of understanding relate to one

another. While it is beyond the scope of this thesis to explore these questions in detail, the account developed in here (Chapters 2-3) makes a case for technological understanding as a key concept for analysing our engagement with technology, and provides a framework for its further theoretical investigation.

The next chapter shifts focus from a theoretical account of technological understanding to its implications for ethics. Discussions about the societal impact of technology constitute a specific form of engagement with technology, and as I will argue, they too build on technological understanding, on a specific type of it in particular.

Chapter 4

Technological Understanding and Ethics

An earlier version of this chapter was published as: de Jong, E. (2026). Functional understanding of quantum technologies is essential to the ethical debate about their impact. Ethics and Society, 20(21). DOI: https://doi.org/10.1007/s11569-026-00501-7

Abstract

As the innovative potential of quantum technologies comes into focus, so too does the need to anticipate their ethical implications. While many voices highlight the importance of ethical engagement, less attention has been paid to the conditions that make such engagement possible. In this chapter, I argue that one key condition is technological understanding: the cognitive skill to recognise how a technology can be used to realise an aim (Chapter 2 and 3). More specifically, I claim that meaningful ethical discussion depends, first and foremost, on understanding a technology's functional capabilities: what it can do and what it can be used for. Presenting such functional understanding as an epistemic requirement helps guide efforts to improve understanding of quantum technologies in support of ethical engagement. These efforts often focus on explaining, in broad strokes, the underlying physics and technical details. While technical insights may indeed sometimes support ethical analysis, they are neither sufficient nor always necessary for grasping the broader societal and ethical implications of a technology. What is indispensable for such discussions is an understanding of its (potential) functions. To foster ethical engagement with quantum technologies, I therefore advocate for promoting functional understanding of these technologies in particular. At first glance, presenting technological understanding as an epistemic requirement for meaningful ethical engagement may seem to raise the bar for participation. However, by decoupling functional understanding from technical expertise, this condition becomes attainable for a broader group, contributing not only to a well-informed but also to a more inclusive ethical debate.

1. Introduction

The investigation of technological understanding in the previous chapters was motivated from the outset by a question about ethics: what kind of understanding of a technology is required for meaningful ethical discussion about it? This chapter takes up that question directly. It connects the account of technological understanding developed in Chapters 2 and 3 to ethics, arguing that ethical engagement with technology is epistemically conditioned and builds on technological understanding.

As was discussed in Chapter 1 (Section 2), quantum technologies represent a new frontier in technological advancement, promising to offer capabilities that surpass many classical physics-based technologies. This potential, often termed 'the quantum advantage', was shown to be associated with a host of opportunities and risks. Consequently, there have been numerous calls to scrutinise the ethical and broader societal aspects of these technologies while they can still be effectively shaped (Coenen & Grunwald, 2017; Inglesant et al., 2021; Hoofnagle & Garfinkel, 2022; Ten Holter et al., 2023; Seskir et al., 2023; Kop et al., 2024a; Gasser, de Jong & Kop, 2024).

Although the calls for ethical engagement with quantum technologies have been loud and numerous, less attention has been given to the conditions that enable such engagement. Selin et al. (2017) argue that effective (public) engagement depends on a range of *capacities* – specifically, the knowledge and skills required for meaningful participation in discussions of science and technology. I argue that one key capacity is 'understanding'. Building on the previous chapters, I propose technological understanding – i.e. recognising how a technology can be used to achieve a practical aim – as a foundational competence necessary for engaging in public discussions about technology. More specifically, I contend that technological understanding is a prerequisite for meaningful ethical debate about the impact of any technology.

However, given quantum's reputation as being 'too complex to understand', this may raise concerns about the possibilities for and accessibility of meaningful ethical discussion. In this chapter, I address these concerns by mobilising a contextual account of understanding – developed in Chapter 3 – that distinguishes different types of understanding, each relevant to a specific context. To foster ethical discussion about quantum technologies, I advocate for promoting a specific kind of technological understanding, focused on the capabilities of quantum technologies and their potential practical deployment, rather than their underlying physics.

First, I explain why an understanding of the technology at hand is crucial to scrutinising its ethical aspects. Second, I summarise the account of technological understanding to clarify what such understanding amounts to, and recapitulate the three main types of technological understanding. Third, I argue that it is, first and foremost, an understanding of quantum technologies' functional capabilities – what they can do and be used for – that is essential for a meaningful ethical discussion. In

doing so, I demonstrate the practical usefulness of the notion of technological understanding and show why distinguishing between different types of technological understanding matters.

Presenting technological understanding as a precondition for ethical discussion raises an epistemic threshold for ethical discussion of emerging technologies like quantum, emphasising the need to understand what a technology can do. But, by stressing the importance of understanding its (potential) functions rather than its underlying physics, this condition also becomes attainable for a wider audience, which fosters an inclusive ethical debate.

2. Understanding as a prerequisite for ethical discussion

2.1 More than knowledge

Imagine a blue device, small enough to hold in your hand, with two buttons labelled "read" and "write". Would you feel sufficiently informed to discuss the ethical implications of this reader-writer? If I told you it was a duplicator, you would still lack essential information – what exactly does it duplicate? Now, suppose I reveal that this device is designed to read electronic data from a car key and write it into a new key. These details are crucial to engage in a well-informed or simply meaningful discussion about ethical concerns raised by the duplicator.

A certain degree of knowledge about a technology seems indispensable for critical thinking about its ethical aspects and broader societal impact, without the need to be a technical expert. Yet, knowledge alone is not enough; it must be combined with the ability to analyse and interpret its implications. This view aligns with Selin et al. (2017) who argue that effective public engagement with science and technology relies on (building) a range of "capacities". These capacities build on both knowledge and analytical skills. Drawing on these considerations, I propose that understanding – as the ability to use knowledge – is a key enabler of meaningful ethical engagement.

In the philosophical literature, understanding has typically been conceived as *the ability to use knowledge* (Reutlinger, Hangleiter & Hartmann, 2018; De Regt, 2017; Grimm, 2012; Wilkenfeld, 2013; Hills, 2016). Some have further specified it as the ability to make counterfactual inferences: to reason about what-would-happen-if scenarios (e.g. De Regt, 2017; Barman et al., 2024; see also Chapter 3). When mapping ethical issues regarding the potential impact of some (future) technology, such inferential abilities play a central role. Debates about the societal impact of technology typically centre on the consequences of using a technology (Swierstra & Rip, 2007, p. 11). This indicates that some form of technological understanding is involved, as the cognitive skill enabling one to foresee the consequences of using an artefact. In other words, ethical engagement can be considered a cognitive task, requiring the ability to

use knowledge about the technology to construct what-if scenarios (*What if the technology materialises in a specific way?*) and make if-then inferences (*If the technology has these capabilities, then those applications are possible.*).

The idea of understanding as an enabling skill for ethical discussion resonates with recent calls to build understanding of quantum technologies (Coenen & Grunwald, 2017, p. 292; Hoofnagle & Garfinkel, 2022; p. xix; Seskir et al., 2023; Grinbaum, 2017; Vermaas, 2017; Quantum Delta Nederland, 2019, p. 40; Rathenau, 2023, p. 3; Roberson, 2023). Vermaas (2017), for example, argues that understanding quantum technologies to "a reasonable degree" is a prerequisite to a societal debate about them. Yet such calls often leave unspecified what this discussion-enabling understanding actually involves and what kind of capacity we need to build. What type or level of understanding of quantum technologies is necessary to consider their ethical aspects?

2.2 Three types of technological understanding

In the previous chapters, understanding in relation to technology was explained as *a cognitive skill that consists in the ability to (recognise how to)[1] use a technology to realise a particular aim*. Applied to, for example, quantum computing, technological understanding thus means: being able to use a quantum computer to achieve a practical goal, e.g. speeding up a specific computational task.

It was also explained that this ability implies that the technology must be intelligible to the user. A technology that is entirely opaque to its user cannot be meaningfully used to achieve an aim. To use a technology successfully, one must be able to reason prospectively about the consequences of operating it: about what will likely happen if it is used in a certain way. This involves grasping the technology's properties and qualities, which in turn can be understood at different levels: technical, practical, or conceptual.

Technological understanding thus requires some insight into how the technology works and what it can do. This is not an all-or-nothing matter, but a cognitive skill that comes in degrees. The required kind and depth of insight depend on the context in which an agent engages with the technology, on the goal being pursued, and the standards of its achievement. In this sense, what counts as (sufficient) technological understanding is context-dependent.

There are three main contexts that involve technological understanding: the context of design, operation and innovation. In each of these contexts – which may overlap in practice[2] – technological understanding is specified in a different way. As was shown in

[1] For this ability can be exercised both conceptually and practically.

[2] As emphasised earlier, the three types of technological understanding are not mutually exclusive: a user may invent a new application, designers may be skilled users, and innovators may participate in

the previous chapters, this yields three contextual *types* of understanding, each foregrounding different aspects of the technology and involving different levels of intelligibility and reasoning. I briefly recapitulate each of these contexts and their associated types of technological understanding.

In a *design context*, where the goal is to conceive and build a technology, technological understanding involves reasoning about the artefact at the level of its inner workings. To design any technology – from a microwave oven to a quantum network – designers need an understanding of an artefact's principles and components in order to design a functioning system. In the case of a quantum computer, this 'technical' understanding consists in recognising how quantum phenomena such as superposition and entanglement can be controlled and manipulated in order to perform computational tasks.

Such design-type or 'technical understanding' is not necessarily required in the *operation context*, where the focus is on practical use. Here, technological understanding requires recognising the direct consequences of interacting with the technology. A driver, for example, must understand how pressing the pedals or using the gear stick affects the car's movement, without needing to comprehend the engine's mechanics. Similarly, 'operational' understanding of a quantum computer involves knowing how to practically operate it and successfully perform tasks with it, rather than understanding its internal physical and technical mechanisms.

In the *innovation context*, the central aim is to invent a new relationship between a technology and a practical aim. While design involves 'thinking up' the artefact itself, innovation – as understood here – entails thinking up its instrumentalisation: determining what a technology can be used for. This includes both repurposing existing technologies – for example, leveraging the capabilities of graphic processing units (GPUs) to train artificial neural networks – and identifying new problems or needs that could be addressed by a yet-to-be-designed technology. Doing so requires recognising a technology's functional capabilities. That is, understanding the principles that determine what kinds of tasks a technology enables, and where its limits lie.

Importantly, this is not the same as superficial technical knowledge; it requires substantive insight into the aims for which the technology may be deployed. For example, recognising the functional capabilities of a large language model (LLM) involves understanding that such models recognise patterns in data and generate text by probabilistically predicting plausible next words. This makes LLMs suitable for some tasks, but unreliable for others.

---

the design process. Distinguishing between them highlights the different ways technological understanding can take shape, foregrounding certain aspects while backgrounding others.

The innovation-type or 'functional' understanding focuses on what the technology *can be expected to do*, rather than how its internal architecture is structured or how exactly it is operated, and thereby enables assessment of its fitness for purpose. In the case of a quantum computer it consists in recognising what it may plausibly be used for (e.g. breaking current encryption methods or speeding up aspects of drug design), as well as what it is not expected to do (e.g. provide an all-purpose computational speed-up over classical computers or automatically 'supercharge' artificial intelligence). I will argue that it is this functional understanding of a technology that is required for a meaningful ethical discussion about its implications.

In the next section, I draw the practical implications of this contextual account and link ethical engagement specifically to the functional type of understanding.

3. A functional understanding of technology

3.1 From physics to functions

Thus far, I have argued that understanding is a key enabling capacity for ethical engagement (Section 2) and proposed technological understanding as the relevant kind (Section 3). Building on this, I will now argue that ethical discussions about quantum technologies – particularly those focused on assessing impact – require, first and foremost, the functional type of technological understanding: understanding what they can do and what they can be used for.

Recall the duplicator. To identify the ethical concerns this machine raises, it is key to understand what it actually *does*. Similarly, in ethical discussions about a quantum computer, what matters is understanding its capabilities and the ways these might be used. Assessing the potential impact of quantum technologies thus requires envisioning their possible applications. This ability aligns primarily with the innovation-type or functional understanding, which entails grasping a technology's capabilities and the range of uses they enable. In short, ethical engagement with quantum technologies requires an understanding of their potential functions.

In practice, however, efforts aimed at fostering public understanding of quantum technologies often tend to focus on promoting a general understanding of their underlying physics. An empirical study by Van de Merbel et al. (2024) found that 81% of Dutch news articles about quantum technologies explain the physics behind quantum technologies. In the same study, the ability to recognise quantum phenomena is taken as an indicator of knowledge about quantum science and technology (Van de Merbel et al., 2024). Another study found that, in 70% of the TEDx talks given by experts about quantum science and technology, one or more quantum phenomena are explained (Meinsma et al., 2023). Introducing quantum technologies to a broader audience, it seems, often means introducing quantum mechanics as well.

This also aligns with my experience over recent years, where introductions to quantum technologies in public-facing and interdisciplinary contexts often focus on explaining the underlying physics. Such introductions are typically not highly technical or mathematically detailed, but instead offer schematic explanations of core quantum phenomena (such as superposition or entanglement) as a way of making quantum technologies intelligible to non-specialist audiences. One plausible explanation for this is the early stage of the field: with its capabilities still largely conceptual and many applications still uncertain, the underlying physics often serves as the primary – or only safe – point of introduction.

This trend is also visible in academic and policy discussions: in this emerging field, many publications on the ethical and societal aspects of quantum technologies include sections on quantum phenomena such as superposition and entanglement (Hoofnagle & Garfinkel, 2022; Kop et al., 2024a; Possati, 2023; Perrier, 2021; Rathenau Instituut, 2023; Vermaas, 2017; Top, 2024). These patterns reveal a 'physics-first approach' to fostering understanding of quantum technologies in which scientific and technical explanations are treated as the primary route to engage people beyond physics communities. This tendency is illustrated by Vermaas' specific call to philosophers of physics "for a renewed effort to make quantum theory understandable" in order to equip a broader group of stakeholders to engage in the societal debate on quantum technologies, suggesting that their expertise uniquely positions them to improve understanding among a wider audience (2017, p. 242).

While there is variation in the degree of technical detail, a physics-first approach consistently foregrounds technical explanations to promote understanding of quantum technologies. The issue here is less with the 'physics' and more with the 'first'. Basic explanations of quantum mechanics can spark public interest (Grinbaum, 2017; Meinsma et al., 2024) and provide a useful backdrop for those who are curious. However, they generally do not make quantum technologies epistemically accessible in a way that enables meaningful ethical discussion. After all, most people do not fully understand how combustion engines, transistors, or GPUs work, yet they can engage in ethical debates about cars, smartphones, and AI. Likewise, explanations of quantum phenomena, quantum circuits, quantum gates, and quantum bits ('qubits') are of limited help when exploring the potential impact of quantum computing: at best, they offer a vague sense of what makes quantum computing different, but they rarely equip people to understand how those differences translate into societal or ethical implications. To explore such implications, one must first develop a sense of the kinds of tasks a quantum computer promises to perform. While those capabilities are ultimately rooted in technical details, it is possible, to a large extent, to develop a functional understanding of them without first acquiring technical understanding (i.e. design-type understanding).

Mentioning 'superposition' in discussions of a quantum computer is, of course, not a crime; it would be odd to omit it entirely. However, focusing too heavily on technical explanations risks misalignment with the aim pursued. Foregrounding the physics aligns more closely with *scientific* than with *technological* understanding (see Chapter 2 for a discussion of their difference). Insofar as it addresses technological understanding, it responds to the technical, design-type, which emphasises how physical phenomena interact with the technology's structure. Often, such technical understanding – focused on a technology's inner workings – is not essential for ethical reflection on its risks, benefits, and broader implications. Moreover, an overemphasis on physics, even in simplified form, may obscure rather than clarify quantum technologies, potentially hindering engagement and meaningful ethical discussions.

Stressing the primacy of functional understanding in enabling ethical engagement does not mean that technical understanding is irrelevant to such discussions. On the contrary, technical insights may substantially support ethical analysis. Some degree of technical understanding can be necessary for recognising functional capabilities and potential applications, and insight into a technology's inner workings may illuminate specific issues, such as those related to direct safety, materials, or dependencies within a technology's stack, that are less visible through other forms of understanding. Functional understanding, however, is indispensable for discussing a wide range of ethical issues and therefore constitutes the core epistemic requirement for ethical engagement. A more technical understanding can complement this, but on its own is insufficient for a meaningful ethical discussion.

To foster meaningful ethical discussions about quantum technologies, especially in relation to their potential impacts, a functional approach to understanding is thus key. Rather than focusing on quantum mechanics, efforts to promote ethical engagement should focus on grasping the possible functions of these technologies in practice. Recent literature appears to be moving in this direction, increasingly favouring functional discussions of quantum technologies over physics-focused explanations (Inglesant et al. 2021; Roberson 2023; Ten Holter et al. 2023; Purohit et al 2024; Root 2025). This is a much-needed development, now grounded in the theory of technological understanding, and deserves wider uptake both within and beyond academia. Importantly, fostering genuine functional understanding requires more than outlining possible applications: it involves enabling people to recognise the functional capabilities that underpin those applications and to grasp the scope of other potential uses.

Finally, note that the functional understanding required for scrutinising a technology's societal impact is essentially of the same type as that involved in innovation. Both centre on grasping a technology's functional capabilities and envisioning its potential applications. The emphasis, however, differs in those contexts: innovation typically requires a deeper or more detailed grasp of those capabilities, while impact assessment

also requires a broader understanding of ethics and the societal context. In the case of early-stage technologies, however, engaging with potential impact requires imagining applications before they materialise, bringing the assessor's understanding closer to that of the innovator.

3.2 Implications for ethical engagement with (quantum) technologies

Presenting functional understanding of quantum technologies as essential for ethical discussions has four implications. First, it suggests that these discussions require more than just knowledge. What is needed is the ability to envision how a technology can be applied. This cognitive skill goes beyond the mere possession of knowledge. Simply transferring factual knowledge about a technology's potential uses is in general not sufficient to enable effective participation in discussions about its societal impact; individuals must be able to imagine and anticipate its possible applications.

Second, it raises *and* lowers the bar for ethical engagement at the same time: it establishes a necessary epistemic condition for engaging in such discussions, while simultaneously making such understanding more accessible by decoupling it from technical, physics-oriented expertise.

Third, adopting a functional approach to fostering understanding of quantum technologies sets agendas: efforts to promote ethical engagement should focus on how these technologies might function in practice rather than on their inner workings. Where the functional capabilities of quantum technologies remain unclear, this points to the need to improve our understanding of what these technologies can and cannot do; and as long as this remains uncertain, ethical engagement should proceed cautiously when addressing questions of impact to avoid speculative overreach (Chapter 5 discusses this further).

Fourth, prioritising functional over technical understanding suggests that those with expertise in quantum mechanics or the inner workings of quantum technologies, despite their deep knowledge, are not automatically well positioned to engage in ethical reflection. They too must acquire a distinct kind of understanding focused on the potential functions and applications of the technology in order to engage with its ethical aspects.

Finally, this argument should not be read as a form of technological determinism. Focusing on functional understanding does not mean accepting the technology as a given and limiting ethical scrutiny to questions of deployment. On the contrary, grasping what a technology can do and may be used for equips us to scrutinise those uses, and to ask the more fundamental question of whether the technology should be developed at all, and if so, under what conditions. Far from narrowing the ethical horizon to questions of how to use a technology well, functional understanding brings prior questions of social desirability and ethical acceptability into view.

4. Conclusion

The innovative potential of quantum technologies requires well-informed ethical discussions to ensure that their development and deployment is not only a technical success but also a net benefit for society and the planet. I have argued that effective ethical engagement is epistemically conditioned: it requires understanding of the technology. I used the previously developed account of technological understanding to clarify what such understanding entails: the ability to recognise how a technology can be used to achieve a practical aim. This perspective provides grounding and guidance for promoting understanding of quantum technologies to support ethical engagement.

Presenting technological understanding as a precondition for meaningful ethical discourse does not imply that ethicists or those involved in ethical discussions need to be technical experts, nor that technical experts are automatically well-positioned to engage in ethical reflection. Distinguishing between different types of technological understanding, I argued that one type in particular – functional understanding – is essential for ethical engagement. What matters is grasping the potential functionalities of quantum technologies to critically reflect on their ethical implications. Without such understanding, ethical discussions risk becoming overly speculative and detached from the realities of the technology.

This claim is one of relevance rather than exclusivity. Other types of understanding – insight into a technology's inner workings or a practical user's perspective – can also illuminate ethically relevant aspects, and ideally do so. The argument of this chapter, however, is that functional understanding is the one most directly relevant, bringing a broad range of ethical issues into view in the first place.

Advancing understanding of quantum technologies' potential functions is therefore a precondition for meaningful ethical discussion about them and their potential impact. While this raises an epistemic threshold for such discussion, it simultaneously makes it more accessible to a wider audience, supporting not only a well-informed but also a more inclusive ethical debate.

The account of technological understanding developed here will resurface in the next part of this thesis, offering a useful perspective on how quantum technologies are currently envisioned and anticipated. The infancy of these technologies necessarily constrains understanding of what they will ultimately be able to do. This constraint actively shapes the discourse, leaving space for assumptions and generating claims that are sometimes more speculative than acknowledged. The next chapters scrutinise this dynamic within the societal and ethical discourse on quantum technologies. In those analyses, I seek to foster reflexivity about its epistemic limits, and develop tools for navigating ethical engagement when understanding is still nascent.

PART II

Part II

Introducing the Ethics Readiness of Technology

How can we engage with the ethical aspects of a technology still shrouded in profound uncertainty? How can we steer between the risks of speculative hype on the one hand, and ethics as an afterthought on the other? This part explores the conditions of 'upstream ethics' – ethics at a very early stage of technological development – and demonstrates how these considerations can guide the ethical study of quantum technologies. In the first part of this thesis, understanding was proposed as a prerequisite for engaging with the ethical aspects of technology. This part takes up the situation in which technological understanding is necessarily limited by the fact that the technology is still in its infancy. Chapter 5 lays the theoretical groundwork for navigating this situation, developing a framework that aligns ethical approaches with a technology's stage of development. Chapter 6 puts this framework to work and approaches the ethical study of quantum computing not by analysing its potential consequences but by investigating the visions that currently shape its development. Chapter 7 addresses one of the ethical issues highlighted in this vision assessment and examines the technosolutionist logic exhibited in the notion of Quantum for Good. Finally, Chapter 8 scrutinises the discourse around the cyber security risk posed by possible future quantum computers and asks under what conditions it may be ethically defensible to treat a still-speculative technology as posing an actual threat that warrants precautionary action. Together, these chapters illuminate the limited technological understanding in the emerging ethical discourse on quantum technologies as an ethical challenge, and propose tools to navigate it.

Chapter 5

Ethics Readiness of Technology

An earlier version of this chapter was published as: de Jong, E. (2025). Ethics readiness of technology: The case for aligning ethical approaches with technological maturity. Science and Engineering Ethics, 31(5), 30. DOI: https://doi.org/10.1007/s11948-025-00556-x

Abstract

The ethics of emerging technologies faces an anticipation dilemma: engaging too early risks overly speculative concerns, while engaging too late may forfeit the chance to shape a technology's trajectory. Despite various methods to address this challenge, no framework exists to assess their suitability across different stages of technological development. This chapter proposes such a framework. I conceptualise two main ethical approaches: outcomes-oriented ethics, which assesses the potential consequences of a technology's materialisation, and meaning-oriented ethics, which examines how (social) meaning is attributed to a technology. I argue that the strengths and limitations of outcomes- and meaning-oriented ethics depend on the uncertainties surrounding a technology, which shift as the technology matures and understanding of its capabilities and implications develops. To capture this evolution, I introduce the concept of *ethics readiness* – the readiness of a technology to undergo detailed ethical scrutiny. Building on the widely known Technology Readiness Levels (TRLs), I propose Ethics Readiness Levels (ERLs) to illustrate how the suitability of ethical approaches evolves with a technology's development. At lower ERLs, where uncertainties are most pronounced, meaning-oriented ethics proves more effective; while at higher ERLs, as impacts become clearer, outcomes-oriented ethics gains relevance. By linking Ethics Readiness to Technology Readiness, this framework underscores that the appropriateness of ethical approaches evolves alongside technological maturity, and helps to ensure scrutiny remains grounded and relevant. Finally, I demonstrate the practical value of this framework by applying it to quantum technologies, showing how attending to their ERLs can guide more effective and appropriately timed ethical engagement.

1. Introduction

The idea that ethics should not be an afterthought is taken by many to be a guiding principle in anticipating ethical issues raised by emerging technologies. The underlying idea is that ethical and broader societal concerns are best addressed as early as possible, before accidents happen or undesirable trends become entrenched, and while the technology is still malleable. Thus, proactive ethical engagement is essential to prevent ethics from arriving 'too late' (Grunwald, 2005, p. 200) when negative consequences may already be manifest, or shaping the design and deployment of a technology becomes significantly harder.

Quantum technologies[1], as a family of emerging technologies, offer a prime opportunity to enact this proactive stance (see Chapter 1 for a brief introduction). Given their potential societal impact, calls have been made for a "responsible approach", emphasising early identification and mitigation of ethical issues (Coenen & Grunwald, 2017; Inglesant et al., 2021; Ten Holter et al., 2023; Seskir et al., 2023; Kop et al., 2024a, 2024b; Gasser et al., 2024). With lessons learned from fields like nanotechnology and AI, the ambition to 'do things right from the outset' has intensified and shapes the emerging debate about the ethics of quantum technologies.

While the early stage of quantum technologies offers the chance to effectively shape their trajectory, it also presents a knowledge problem. There is still considerable uncertainty about whether, when, and how promising technologies like quantum computing and quantum communication will materialise. Given this uncertainty and the inherent complexity and unpredictability of technological innovation generally, the anticipation imperative may force early-stage ethical discussion to rely on a range of assumptions. Facing the risk of becoming overly speculative, ethics thus may also come in 'too early'. How, then, can we balance the timing of ethical considerations for emerging technologies like quantum – avoiding coming in too soon or too late?

Over the last few decades, a wide range of methods has been developed for the ethical study of early-stage technologies.[2] There has been ongoing debate about the strengths and weaknesses of these methods, with none emerging as dominant (Reijers et al., 2018). An overarching framework for evaluating the applicability or *appropriateness* of these methods – understood here as their suitability given contextual constraints and goals – remains lacking. This chapter seeks to fill this gap by providing such a framework, one that links the appropriateness of ethical methods to a technology's stage of development. In particular, it addresses the question: How can we align our

[1] This thesis focuses exclusively on second-generation quantum technologies, which harness quantum mechanical phenomena for emerging applications such as enhanced sensing, advanced simulation and computation, and novel communication systems. By contrast, first-generation quantum technologies include well-established innovations like transistors, lasers, and magnetic resonance imaging (MRI).

[2] Reijers et al. (2018) identify 35 different methods for practicing ethics in research and innovation.

approach to ethics with the stage of technological development to ensure timely and meaningful ethical consideration? In other words, how can we ensure that ethical considerations in relation to emerging technology are neither premature nor overdue?

To answer this question, I distinguish between two main ethical approaches: *outcomes-oriented ethics*, which assesses the potential consequences of a technology's materialisation, and *meaning-oriented ethics*, which examines how (social) meaning is attributed to a technology (Section 2). I then take this analysis further to argue that the appropriateness of these approaches depends on the uncertainties surrounding a technology, which shift as it matures. To capture this evolution, I introduce the concept of *ethics readiness* – the readiness of a technology to undergo detailed ethical scrutiny (Section 3). Building on the widely known Technology Readiness Levels (TRLs), I propose Ethics Readiness Levels (ERLs), which indicate the types of ethical questions that are appropriate at different stages of a technology's development. The resulting framework helps to assess the suitability of ethical approaches based on a technology's maturity and can guide ethics in practice. Finally, to explore the practical implications of the framework for guiding ethical discussions of emerging technology, I apply it tentatively to the case of quantum technologies (Section 4).

2. Outcomes- and meaning-oriented approaches to ethics of technology

In this section, I explore the landscape of methods in the ethics of emerging technologies. I begin by identifying the central challenge in this field: addressing the epistemological issue of uncertainty and ignorance (2.1). Next, from the literature I distil two primary approaches to the ethics of emerging technologies: those focused on anticipating future outcomes (2.2) and those centred on understanding the present meanings attributed to a technology (2.3). Building on this distinction, I argue that the appropriateness of these approaches varies depending on the specific stage of technological development (2.4). This analysis lays the groundwork for the concept of 'ethics readiness' which is further developed in Section 3.

2.1 Upstream ethics and the problem of uncertainty

The ethical study of emerging technologies and their impacts – commonly referred to as the ethics of new and emerging science and technology (NEST) or simply NEST-ethics (Swierstra & Rip, 2007) – is inherently anticipatory. To understand this anticipatory character, it can be helpful to imagine technological innovation as a 'stream': to avert or mitigate undesirable outcomes and promote desirable ones 'downstream', it is essential to make informed decisions 'upstream', early in the innovation process. This *upstream ethics*, as I will refer to the ethics of early-stage technologies, aligns with the broader paradigm of responsible research and innovation (RRI), which is generally geared towards anticipation and integrating societal

considerations throughout the research and innovation process (Ryan & Blok, 2023; Genus & Stirling, 2018; Wickson & Carew, 2014).

Anticipating ethical concerns highlights a fundamental challenge to upstream ethics and RRI more generally: navigating an unknown future (Lucivero et al., 2011; Sollie, 2007). As Swierstra and Rip (2007, p. 6) aptly state: "Working with a novelty necessarily means venturing into the unknown." This uncertainty hallmarks all NEST-ethics, ranging from debates about Artificial Intelligence and quantum computers, to lab-grown meat, cloning, and genetic engineering. The complexity and unpredictability of technological innovation implies uncertainty about a technology's trajectory and ignorance about its consequences (Ferrari, 2010). In this sense, anticipating ethical concerns is 'doubly fictional' (Rip & te Kulve, 2008): grappling with yet-to-materialise technologies and their yet-to-manifest implications.

This 'problem of uncertainty' is rooted in the very character of innovation itself, which is inherently geared towards actively seeking novelty (Steen et al., 2021; Nordmann, 2014). This aligns with Lucivero et al. (2011), who state that the object of ethical assessment of emerging technology is, by definition, "elusive" (p. 129). While uncertainty and unpredictability are intrinsic to innovation, this elusiveness is most pronounced during the earliest stages of research and development.

This epistemic problem can be reframed as a problem of technological understanding (see Chapters 2–4). When a technology is still in a conceptual or early experimental phase, understanding of what it can do and may be used for is necessarily limited. If such functional understanding is key for meaningful ethical engagement – as argued in Chapter 4 – then limited understanding at the earliest stages of development translates directly into constrained ethical engagement: the questions that can meaningfully be asked are limited by what is known.

The counterpart of this early-stage knowledge problem is a late-stage power problem: as the technology matures and becomes embedded in society, the capacity to effectively adapt its design and deployment diminishes. This tension is captured by Collingridge's (1980) well-known dilemma of social control of technology, which highlights the trade-off between early-stage uncertainty and late-stage entrenchment. This dilemma has an ethical variant, with ethics either coming in 'too soon', "being too speculative to be reliable", or coming in 'too late' to effectively shape the technology's impacts (Kudina & Verbeek, 2019).

One often-advocated response to the dilemma of social and ethical control focuses on fostering "technological flexibility" or corrigibility (Collingridge, 1980; Joly, 2015; Genus & Stirling, 2018). This involves designing technology to be adaptable and responsive to feedback as new social or ethical concerns arise, in theory allowing society to retain a degree of control. Yet realising such flexibility, and ensuring that it adequately addresses ethical concerns, presents challenges. First, the design of flexible

technologies itself requires foresight about potential ethical issues, which revisits the knowledge problem. Second, some ethical issues, particularly those warranting precautionary action, cannot be fully mitigated by flexible design – if meaningful adaptability is possible at all. For instance, the cyber security risks associated with largescale quantum computing cannot be fully mitigated through adaptability alone; the technology's core capabilities could undermine current security protocols. Technological flexibility is thus not a panacea, and it risks backgrounding the anticipatory ethics necessary for emerging technologies.

Anticipating a late-stage power problem – where the opportunity to shape a technology's trajectory is lost due to accumulated, practically irreversible decisions – inherently involves addressing the early-stage knowledge problem and thus finding ways to responsibly navigate uncertainty. Over the past two decades, numerous methods have been proposed to address this challenge, with a notable increase in the last ten years (Reijers et al., 2018). These methods – commonly referred to as "ex ante methods" – focus on engaging with ethical issues before technologies have materialised, at the earliest stages of research and innovation (Reijers et al., 2018; Urueña, 2024). In the following sections, I distinguish between two primary approaches within such upstream ethics, each offering a distinct strategy for addressing the uncertainties inherent in early-stage ethics.

2.2 Outcomes-oriented ethics: engaging with possible futures

One strategy to engage with unknown futures is to envision possible futures and take these as the object of ethical study and debate. Prominent examples include ETICA (EThical Issues of emerging iCt Applications) (Stahl & Flick, 2010), ethical technology assessment (Palm & Hansson, 2006), techno-moral scenarios (Boenink et al., 2010), and anticipatory technology ethics (Brey, 2012).[3] Methods of this sort all employ a type of *ethical foresight analysis* (Floridi & Strait, 2020): they aim to envision a future technology and their possible implications. These possible futures are then taken as the object of ethical scrutiny.

[3] Stahl and Flick (2010) position the ETICA approach within the broader field of foresight activities, noting that "the findings of the project have the character of outlining possible futures with a view to identifying futures that are desirable and that can be influenced by present action" (p. 65). Similarly, Palm and Hansson (2006) describe ethical technology assessment as "a tool for identifying adverse effects of new technologies at an early stage" (p. 543). Swierstra, Stemerding, and Boenink (2010) introduce the development of techno-moral scenarios – which depict a broad range of potential future consequences arising from the interaction between technology and morality – as a method to stimulate ethical reflection (p. 120). Brey (2012) characterises anticipatory technology ethics as the "anticipation of possible future devices, applications, and social consequences" through "methodologically sound forecasting" (p. 1).

Yet, as multiple scholars have noted – including most of the authors of the approaches mentioned above – early ethical reflection can become speculative when uncertainty about a technology's development and applications remains high (Nordmann, 2007; Swierstra & Rip, 2007; Nordmann & Rip, 2009; Brey, 2012; Ferrari, 2010; Grunwald, 2017; Reijers et al., 2018). This critique of speculative ethics has most prominently been voiced by Nordmann (2007, 2014). Central to Nordmann's critique is that ethics excessively leaps ahead of science. By relying on too many assumptions, ethics "foreshortens the conditional", treating hypothetical scenarios as presenting actual ethical issues (Nordmann, 2007, p. 32). Such "thinking ahead too much" (Gilbert & Goddard, 2014) risks an arbitrary focus, potentially diverting intellectual resources to unlikely scenarios while neglecting pressing ethical issues.

The critique of speculative ethics raises questions about framing the potential ramifications of emerging technologies as 'risks' – commonly defined as a probability multiplied by impact – when the probabilities involved are sheer unknowns (see Chapter 8). Efforts to mitigate speculative risks might divert attention from other, more pressing, issues. However, anticipating risks is fundamental to ethics' role in shaping technology, thus requiring it to leap ahead of scientific and technological developments. Importantly, when the expected impact of a technology is potentially vast, even low-probability scenarios might warrant serious ethical consideration (Price, 2024). I will return to this issue at the end of Section 3.

Acknowledging the need to think ahead, scholars like Roache (2008) and Urueña (2022, 2023) argue that ethical debate should not be confined to current technology: they defend the importance and possibility of anticipating possible future scenarios. Likewise, Grunwald (2010) and Selin (2014) hold that engaging with possible futures is a useful reflexive heuristics. Even critics of speculative ethics admit that controlled or 'responsible' speculation is essential to ethical enquiry (Nordmann & Rip, 2009). Thus, any upstream ethics must engage with possible futures. The challenge is how to do so effectively while carefully avoiding the pitfalls of speculation.

Addressing this challenge calls for a closer look at what critiques of speculative ethics are actually targeting. These critiques often focus on the predictive mode of anticipation, which centres on *future outcomes*. NEST-ethics has indeed been associated with consequentialist patterns of reasoning that evaluate actions based on their (potential) results rather than the actions themselves or the intentions behind them (Swierstra & Rip, 2007; Ferrari, 2013; Grunwald, 2020): "The new and emerging technology is deemed desirable, or not, because its consequences are desirable, or not." (Swierstra & Rip, 2007, p. 11) However, while this consequences-focused approach aligns with consequentialist reasoning, it does not necessarily imply strict consequentialism, which locates all normative value solely in outcomes. For this reason,

I refer to the consequentialist tendency in upstream ethics more broadly as 'outcomes oriented.'

The outcomes-oriented approach to ethics, or in short 'outcomes-oriented ethics', typically focuses on the (ethical) implications and consequences of a technology, which can occur at multiple levels. Taking Artificial Intelligence (AI) as an example, this approach would focus ethical enquiry on the implications of design choices in models, and on potential uses, as well as the direct effects of that use. Additionally, this focus on outcomes could expand to more indirect societal impacts, such as changes in social dynamics, deskilling, and privacy concerns.

Several authors argue that an ethics focusing on consequences or outcomes faces challenges due to the inherent unpredictability of innovation (Grinbaum & Groves, 2013; Nordmann, 2007; Sand, 2018; Steen et al., 2021; Vallor, 2016). An emphasis on potential outcomes assumes a level of foresight that clashes with the uncertainty of early technological development. NEST-ethics thus risks excessive speculation when assessing possible outcomes at a nascent stage. In essence, the critique of speculative ethics targets attempts to conduct 'downstream ethics' – focused on assessing specific consequences of a concrete technology – when still at a very 'upstream' point in its development.

Beyond predictive anticipation, there are alternative modes of engaging with technological futures. In responsible innovation literature, authors like Roache (2008) and Urueña (2022) highlight the value of explorative, strategic, and critical-hermeneutic modes of anticipation. Shifting from predictive to hermeneutic anticipation at the earliest stages of innovation suggests moving away from a focus on speculative outcomes and towards examining how, by whom, and under what assumptions these outcomes are envisioned and given significance. This alternative perspective on anticipating and engaging with possible futures introduces a second approach to ethics, which the next section examines.

2.3 Meaning-oriented ethics: assigning meaning to future technologies

Besides the outcomes-oriented approach, a second approach has emerged within the ethics of emerging technology. This approach seeks to mitigate 'speculative excess' by grounding ethical enquiry more firmly in the present. Some of them aim to reduce speculation by basing ethical enquiry in concrete practices. Examples are real-time technology assessment (Guston & Sarewitz, 2020), and experimental ethics (van de Poel, 2016). Yet, because these methods depend on a technology's actual level of development, they may face limitations in early-stage anticipation.

Other methods that seek to ground ethics in the present focus on "the future as it exists already" (Nordmann & Grunwald, 2023). Since emerging technologies are emerging, they predominantly exist in the form of visions, promises, and expectations

(Borup et al., 2006; Lucivero et al., 2011). These "future-oriented abstractions" (Borup et al., 2006) capture – individual or collective – ideas about how the future might look like, "often expressed in the semantic of intentions, goals, hopes or proposals." (Lucivero et al., 2011, p. 131). Also, these abstractions tend to focus on the generic qualities of a technology, rather than specific artefacts and applications.

According to Grunwald (2017), envisioning 'technology futures' – imaginaries of what society might look like when a specific technology materialises – entails assigning *meaning* to a technology; the technology becomes something to appraise. Expectations and visions inform us about the perceived desirability and acceptability of technologies. These assignments of meaning are value-laden, normative acts and hence, can be subjected to ethical reflection. In fact, Grunwald argues that doing so is important because meaning-assignment activities – the way we think and talk about the future – set the stage and continue to influence the subsequent debate (Grunwald, 2023; see also Van der Burg, 2014). Or as Lucivero et al. (2011) put it: "[T]here is a self-referential loop between the present and the future: the way in which we *describe* the future will determine *how* the future will be." If we seek to study an emerging technology, we thus need to study how it is being perceived and ethically assess the meanings attributed to it.

Several methods have been proposed for the ethical study of meaning assignment by critical reflection on "sociotechnical imaginaries" (Jasanoff, 2020). These include vision assessment (Schneider et al., 2021, Nordmann, 2007, 2014; Grunwald, 2004, 2014, 2020), reflection on metaphysical-research programmes (Ferrari, 2010), and hermeneutic technology assessment (Nordmann & Grunwald, 2023).[4] What connects these methods, is the idea that critical examination of sociotechnical imaginaries can help identifying the values, assumptions, and power dynamics that shape technological innovation (Urueña, 2022). It also brings into view *who* shapes these visions and *for whom*. In other words, critical examination of visions can reveal the 'impact-makers,' those with the power to shape visions and a technology's trajectory, and what may be called 'impact-takers', those who lack that power but are likely to experience the impact of a technology.

I will collectively refer to the ethical methods focused on meaning assignment and fundamental questions about the sociotechnical imaginaries surrounding a technology as the 'meaning-oriented' approach to ethics or, in short, 'meaning-oriented ethics.'

[4] Vision assessment examines the future narratives introduced and promoted by specific actors, networks, or organisations for strategic purposes (Schneider et al., 2021, p. 1091). Hermeneutic technology assessment takes a similar approach by reflecting on the meanings people associate with a given technology, but extends the scope beyond visions to include a broader range of expectations, hopes, and fears across a wider group of stakeholders. enquiry into metaphysical research programmes involves ethical reflection on the cultural, historical, and political contexts in which technological development occurs (Ferrari, 2010, p. 38).

Meaning-oriented ethics aims to facilitate ethical discussion at the early stages of a technological project, while reducing speculation about how the future will look like. The analysis of meaning assignment is not hindered by the problem of uncertainty in the same way outcomes-oriented ethics is. Although 'techno-visionary futures' (Grunwald, 2023), or in short 'technovisions', involve images of what the future with a specific technology may look like, it is not the results themselves that are the object of ethical scrutiny, but rather the act of envisaging them. Nonetheless, ethical study of visions, promises, and expectations necessarily includes the assessment of the quality of the claims that constitute them – not only taking technological feasibility into account, but also the perceived social usability and desirability (Lucivero et al., 2011). In that way, the probability and plausibility of technology futures become the object of ethical scrutiny, and responsibly dealing with uncertainty takes shape as fostering 'responsible representation' (Nordmann & Rip, 2009).

Besides this hermeneutic approach, which focuses on the process of envisioning futures, meaning assignment can also be studied from a conceptual point of view. Lunshof and Rijssenbeek (2024) consider the earliest phases of technology development 'pre-ethical'. At this stage of fundamental research and early development, they consider the technology too premature to be subjected to actual ethical analysis. To prepare for such ethical analysis at later stages, they argue that it is the task of early ethical engagement to reflect on the meaning of central concepts and on how the technology in question may challenge or disrupt them. Instead of labelling such conceptual analysis as pre-ethical – as Lunshof and Rijssenbeek do – I assert that this kind of research can be subsumed under meaning-oriented ethics since it focuses on how we attribute meaning to a technology through (revising existing or inventing new) concepts.

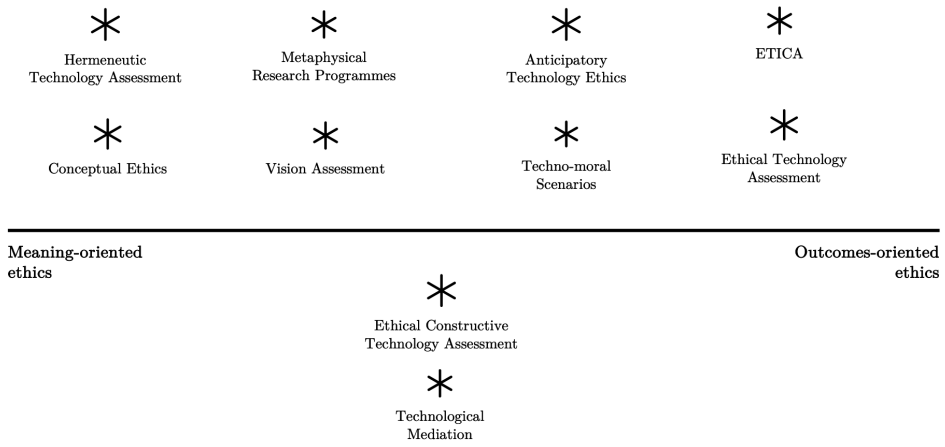
Revisiting the example of AI, meaning-oriented ethics would focus on examining the values that underpin its development, questioning the desirability of the envisioned futures, and critically reflecting on how the deployment of AI may affect existing power structures and vice versa. It could also involve exploring how AI challenges established definitions of values and disrupts fundamental concepts, like autonomy, control, and humanity itself. This type of ethical analysis thus focuses more on structural issues associated with a future technology, instead of the implications of specific applications.

Before turning to the question of when each approach is most appropriately applied, it is important to emphasise that the distinction between outcomes-oriented and meaning-oriented ethics is conceptual rather than categorical. In practice, the two approaches should be seen as points on a spectrum, with methods positioned according to their primary focus – whether on the 'input' of the innovation process (the ideas, expectations, and visions shaping a technology and its trajectory) or on the 'output' (its implications, consequences, and impact). Many methods gravitate to either one of the poles, but some fall somewhere in between. For example, ethical constructive

technology assessment (Kiran et al., 2015) and the technological mediation approach (Kudina & Verbeek, 2019) exemplify this: they aim to anticipate ethical implications while also attending to the processes through which technologies acquire meaning. For analytic clarity, these hybrid approaches are not further discussed here, yet they remain important to consider as transitional ethical strategies relevant during stages of technological development when a technology begins to take on a more concrete and defined shape.

Figure 1 gives a non-exhaustive overview of the outcomes- and meaning-oriented approaches to ethics.

Figure 1. Methods within the outcomes- and meaning-oriented approaches to the ethics of emerging technology (not exhaustive)



3. Evaluating ethical approaches: adding a temporal dimension

The outcomes- and meaning-oriented approaches to ethics deal with uncertainty in different ways. Outcomes-oriented ethics tackles uncertainty by speculating about potential impacts, creating an object of anticipation. However, this can lead to speculative overreach, especially in the early stages. In contrast, meaning-oriented ethics focuses on grounding ethical analysis in the present, examining the visions, promises, and assumptions that shape how technologies are imagined. This focus brings the analysis closer to the social and cultural contexts in which these imaginaries are formed, offering a more immediate connection to the present. As technologies mature, however, the urgency of assessing potential impacts grows, highlighting the limitations of meaning-oriented ethics.

The strengths and limitations of both approaches are closely tied to timing. As technologies mature, the epistemic conditions change, altering their suitability and relevance. Thus, their appropriateness or meaningful applicability is highly dependent on the specific stage of technological development. Urueña (2024) touches on a similar idea by linking different "modes of anticipation" to variables such as the timing of anticipatory engagement (ex-ante, ex-dure, ex-post). However, the notion that ethical approaches should adapt alongside technological development remains underexplored. In discussions of upstream or ex-ante ethics (Brey, 2012; Floridi & Strait, 2020; Reijers et al., 2018), the term 'emerging' often lumps together stages of research and development that differ in their epistemic conditions, while these differences have implications for ethics. Thus, the ethical study of emerging technologies would benefit from incorporating this temporal dimension into the frameworks for assessing the appropriateness of different approaches.

Similarly, a technology's stage of development shapes what *level of analysis* is appropriate. While useful distinctions have been made – for example, Brey's (2012) focus on technology, artefacts, and applications, and Ryan et al.'s (2024) differentiation between the microlevel (artefact), mesolevel (sociopolitical structures), and macrolevel (ontological considerations) – the relevance of each level in relation to the stage of technological development often remains implicit. However, it is reasonable to assume that the stage of development significantly impacts the relevance of each level. For example, analysing artefacts and applications typically assumes a more advanced stage of development, while earlier stages may require broader considerations, such as those at the technological level.

The framework I propose in the next section introduces a temporal dimension into the evaluation of ethical approaches and links this to the appropriate level of analysis. I argue that different stages of technological development allow for different kinds and levels of ethical discussion. Using the Technology Readiness Level (TRL) scale, I link the relevance of outcomes- and meaning-oriented ethics to specific levels of maturity, aligning ethical scrutiny with the actual stage of technological development. Incorporating this temporal dimension into the evaluation of ethical approaches aligns with Collingridge's (1980) incrementalist approach to the dilemma of control, which stresses that the assessment of technology should evolve alongside its development.

4. The ethics readiness of technology and its implications for ethics

This section develops a framework that explicitly and systematically links the appropriateness of outcomes- and meaning-oriented ethics to a technology's stage of development. It begins by discussing the Technology Readiness Level (TRL) scale and its societal and ethical dimensions (3.1). Drawing on this scale and the discussions in Section 2, I introduce the concept of 'ethics readiness' as a measure of a technology's

preparedness for detailed ethical scrutiny (3.2). Next, I link ethics readiness with the TRL scale, resulting in the Ethics Readiness Level (ERL) scale (3.3). This scale highlights how the appropriateness of ethical approaches depends on a technology's maturity.

4.1 Technology readiness and societal readiness

Since a technology's stage of development determines the appropriate ethical approach, identifying that stage is an important step in ethical analysis. The Technology Readiness Level (TRL) scale provides a concrete framework for distinguishing between different stages. Originally developed by NASA, this framework assesses the maturity level of a specific technology on a scale that defines 'readiness' in relation to market viability. The central idea is that typically, new technologies go through various stages of the TRL scale in their life cycle. The TRL framework is widely adopted as a tool for decision-making, for example by the European Commission (EC) in the Horizon Europe funding programmes. In the EC's version, the TRL scale comprises nine levels of maturation, which can be categorised into four phases (APRE & CDTI, n.d.; Netherlands Enterprise Agency, 2022):

- *Discovery Phase* (TRL 1–3): This exploratory phase includes fundamental research, applied research, and the creation of a proof-of-concept. By TRL 3, the technology is considered scientifically feasible.
- *Development Phase* (TRL 4–6): Encompassing testing, validation, and prototype demonstration, this phase ensures the technology is feasible and reliable from a technological standpoint by TRL 6.
- *Demonstration Phase* (TRL 7–8): This phase involves demonstrating the technology in its operational environment and finalising it for implementation. At TRL 8, the technology is both technically and commercially ready.
- *Deployment Phase* (TRL 9): The final phase, where the technology is scaled and introduced to the market, marking its readiness as a commercial product or service.

The TRL scale predominantly focuses on readiness from a technological perspective, assessing whether *the technology* is ready for deployment in society ('technological feasibility'). The framework does not explicitly and systematically account for whether *society* is ready for the uptake of the technology at hand. However, to effectively and responsibly embed a technology in society, we must consider society's 'readiness' as well. This includes regulatory frameworks, broader sociotechnical ecosystems, and attitudes towards the technology (WRR, 2021; de Jong, 2022). Neglecting society's readiness may obstruct successful uptake and lead to resistance, governance problems, and broader issues of non-alignment with public values.

To gauge societal readiness, Bernstein et al. (2022) developed the Societal Readiness Thinking Tool. The tool is based on the idea that societal readiness is an iterative process of addressing societal concerns during the technology life cycle. These societal concerns are categorised in five dimensions: public engagement, open access, science education, gender, ethics, and governance. By presenting sets of guiding questions for different phases of the TRL scale, the tool seeks to integrate thinking about societal concerns at different stages of the innovation process.

Similarly, Bruno et al. (2020) have proposed an extension of the TRL scale in three directions, resulting in scales for Legal, Organisational and Societal Readiness Levels. Just like technological innovation itself, Bruno et al. argue that compatibility with legal frameworks, organisational structures and societal needs also progresses through various stages of maturity. In their framework, the levels of legal, organisational and societal readiness mirror the Technology Readiness Levels, resulting in nine stages. These stages are primarily explained in terms of 'readiness to adopt' a particular technology, from a legal, organisational and societal point of view. In other words, the focus of this framework is on identifying legal, organisational, and societal (or social) 'enablers and barriers' for the take-up of a particular technology.

Both frameworks can be seen as useful extensions that expand and contextualise the original Technology Readiness Level (TRL) scale, highlighting that high technology readiness does not necessarily imply societal readiness. They broaden the range of considerations included in assessing a technology's readiness for deployment in society. Notably, both frameworks incorporate ethical dimensions: Bernstein et al. (2022) identify ethics as one of the 'dimensions' of societal readiness, while Bruno et al. (2020) include ethical considerations within the scope of legal compliance. These contributions support the idea that ethical discussions evolve and change in character alongside progress on the TRL scale.[5]

However, the idea that different stages of technology development allow for – and even require – different approaches to ethics, remains implicit in these frameworks. Building on the previous discussion about the suitability of specific ethical approaches at various phases of technological development, there is a strong case for explicitly linking the type or 'mode' of ethical engagement to the TRL scale. Beyond its current function of guiding innovation policies and serving as a decision-making tool, the TRL scale could thus also help to determine the appropriate ethical approach for emerging technologies, such as quantum technology.

[5] Note: Advancing through higher TRLs does not imply that a technology becomes more ethically acceptable. Rather, the point is that the nature of the ethical questions it affords tends to shift as the technology matures.

4.2 Ethics readiness

The proposals by Bernstein et al. (2022) and Bruno et al. (2020) to complement the TRL scale with a societal readiness scale lay the groundwork for viewing NEST-ethics as evolving alongside technological development. To capture this, I introduce the concept of *ethics readiness*[6].

I define 'ethics readiness' as the extent to which a technology is prepared to undergo detailed ethical scrutiny. While ethical considerations are relevant from the outset of technological innovation, it is important to distinguish between the general need for ethics and the appropriateness of specific ethical approaches at different stages of technological development. Ethics cannot come 'too early' in a general sense, but certain types of ethical questions may be premature at specific stages.

The ethics readiness of a technology evolves along its technological development. The ethical analysis it supports thus becomes increasingly detailed and specific over time. In the early phases, particularly during the conceptual stage, ethics readiness is typically low due to high uncertainties and a lack of concrete information. As a result, the technology is not yet ready for detailed ethical scrutiny. However, this early stage allows for more abstract ethical analysis, such as that associated with meaning-oriented ethics. As discussed in Section 2.3, meaning-oriented ethics examines the broader visions, promises, and assumptions shaping a technology's trajectory. While this approach rigorously addresses systemic and abstract concerns, it is less granular in addressing specific features of the technology itself.

In contrast, outcomes-oriented ethics focuses on specific features, applications, and impacts of the technology. This approach inherently involves more detailed ethical scrutiny, as it examines concrete aspects of the technology and their potential consequences (see Section 2.2). As the technology matures and uncertainties decrease, its ethics readiness increases, enabling more detailed and specific analyses typically associated with outcomes-oriented ethics.

The idea of a technology's evolving ethics readiness can be further understood through the lens of *critical-reflective affordances*. In a recent paper, Urueña (2024) introduces this term to describe the opportunities that certain modes of anticipation provide for critical reflection on future technologies. These affordances determine the kinds of ethical questions that different frameworks of anticipation allow us to ask, with some modes better suited to certain stages of technological development than others.

[6] I use 'ethics readiness' instead of 'ethical readiness' because the latter may suggest that a technology is ethically ready in the sense of aligned with ethical values. Ethics readiness, on the other hand, encompasses the readiness of a technology for ethics.

Applied to technology[7], critical-reflective affordances describe the *ethical questions and discussions* afforded by a technology at a given stage of development. Each stage of a technology's development offers different entry points for ethical reflection. These entry points are the affordances for critical reflection: the possibilities for scrutiny shaped by the technology's state of development. As technologies mature, these affordances evolve, driven by increasing availability of concrete information and a narrowing scope of uncertainties. At early stages, when a technology is still an idea, it affords reflection on its potential purpose, societal implications, and alignment with values. As it evolves into a working prototype, it affords consideration of design choices, applications, and potential impacts on users or society. This understanding adds depth to the concept of ethics readiness: it reflects the type and specificity of ethical questions a technology affords at specific stages of its development.

Although ethics readiness is conceptually distinct from societal readiness, the two are meaningfully related. Ethics readiness indicates the types of ethical questions most appropriate at a given stage of technological development, and these questions can, in turn, inform broader discussions about society's readiness to engage with or adopt a technology. In this way, the concept of ethics readiness complements existing frameworks for assessing societal readiness.

4.3 Pairing technology readiness with ethics readiness

To assess a technology's ethics readiness, I propose the Ethics Readiness Level (ERL) scale, aligning with the Technology Readiness Level (TRL) scale. This means that a low TRL typically corresponds to a low ERL, while a high TRL corresponds to a high ERL. The idea here is that as a technology matures, its ethical implications become more concrete, increasing its ERL, i.e. its readiness for detailed ethical scrutiny. (This should not be conflated with its ethical acceptability or social desirability.)

At lower TRLs, high uncertainty and limited information result in a lower ERL, constraining the scope for meaningful ethical debate. Conversely, at higher TRLs, empirical data on the technology's effects becomes available, enabling detailed ethical scrutiny and supporting a broader range of methods.

The resulting framework, which systematically links Technology Readiness Levels (TRLs) to Ethics Readiness Levels (ERLs), can help guide the most appropriate approach of ethics for a specific emerging technology. Before it can do so, the framework must first define the readiness levels at which a technology is considered 'emerging'. While there are more detailed definitions of 'emerging technology' (e.g. Rotolo et al., 2015), the term is often used more generally in NEST-ethics to describe technologies

[7] While Urueña applies this concept to the qualities of anticipation frameworks, he also suggests that it can be applied to technologies themselves (2024, p. 8).

that are still in development and not yet widely available, with their (ethical) impact expected to unfold in the future (Brey, 2017; Reijers et al., 2018). I will use this broader definition of 'new and emerging technologies' and identify it with a specific range of TRLs.

As I discussed in Section 2, emerging technologies are typically understood as being in the early stages of development, characterised by significant uncertainty and novel possibilities. These characteristics are most pronounced at lower TRLs (1–3), where technologies remain in conceptual or early prototype phases, and mid TRLs (4–6), where viability is demonstrated, but full realisation is still underway. Thus, I propose that emerging technologies be defined as those at TRLs 1–6, where uncertainty is fundamental and broad in scope – acknowledging that the nature of this uncertainty shifts as a technology progresses from early-stage (TRL 1–3) to mid-stage (TRL 4–6). By the time technologies reach higher TRLs (7–9), they may be considered 'new' but have moved beyond the phase of emergence, as they are closer to full deployment and societal integration.

At lower TRLs (1–3), uncertainty about whether, how, and when a technology will materialise dominates. This typically leads to a low ERL. As discussed in Section 2.4, this level of uncertainty diminishes the effectiveness of the outcomes-oriented ethical approach, which relies on anticipating specific impacts or consequences. In contrast, meaning-oriented ethics is less hindered by these unknowns, making it a more valid and practical approach at the earliest stages of technological development. In other words, the technology's critical-reflective affordances at this stage align best with meaning-oriented ethics. This approach often implies an analytical focus on the broader level of the technology itself, instead of on specific designs and applications.

At mid TRLs (4–6), while the technology is still evolving and not yet fully realised, its technological viability has typically been demonstrated, and there is growing information about specific devices and potential applications. This raises the ERL and broadens the scope of ethical enquiry. As the technology starts to take more specific shape, it becomes both feasible and necessary to anticipate consequences. As a result, at this mid ERL methods within the outcomes-oriented approach gain relevance. This transition is accompanied by a shift in the level of analysis: from the broader focus on technology as a whole to a more specific consideration of artefacts and their applications. While the meaning-oriented approach remains valuable, outcomes-oriented methods increasingly complement it, reflecting the evolving affordances of the technology.

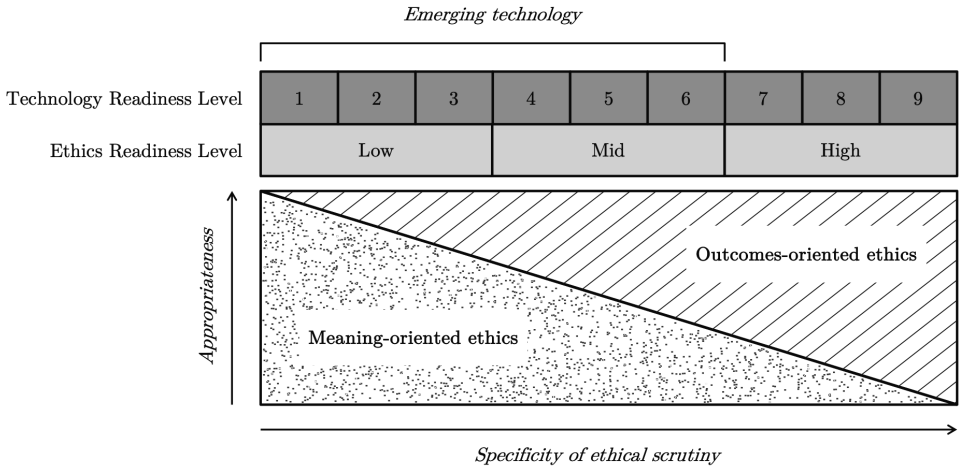
At higher TRLs (7–9), where the technology is practically deployed, its ERL is typically high. At this stage, the relevance and effectiveness of meaning-oriented methods diminish, while the need for outcomes-oriented methods becomes more pressing. Real-world deployment generates real-world impacts, which increases the need to anticipate

and address ethical consequences in a concrete, outcomes-based manner. Moreover, the concreteness of the technology supports increasingly detailed ethical questions.

In the middle stages of the TRL scale, as a technology moves from concept to viable product, more hybrid ethical approaches may be especially appropriate, as they attend to both the processes of meaning-making and the implications of the technology's real-world application (see the discussion at the end of Section 2.3).

Figure 2 summarises the ERL-framework.

Figure 2. Ethics Readiness Level (ERL) scale: The effectiveness and level of specificity of ethical approaches evolves along with technological maturity



This framework offers a broad guide for aligning the ethical approach with the specific stage of technological development. However, it is intended as a starting point rather than a strict rule. While it captures a typical correlation between technology readiness and ethics readiness, it is open to the fact that ethics is not always linear or bound to technical timelines; there may be cases where early-stage technologies warrant ethical questions more commonly associated with later stages.

One factor that may justify early outcomes-oriented ethics might be the severity of potential impact: if the consequences of a given technology could be catastrophic, then the anticipated impact may compensate for its low technology readiness and warrant an outcomes-oriented ethical approach. This will be further illustrated in the next section, through the example of anticipated cyber security breaches enabled by sufficiently large quantum computers. (See also Chapter 8 for a discussion of this case).

Still, even when an early-stage technology has relatively specific contours, draws on prior ethical debates, or presents potentially significant impacts, some indication of

technical feasibility remains necessary to ground ethical discussion and ensure that the questions raised connect to what the technology can actually do. The connection between ERLs and TRLs should be understood as a rule of thumb: ethics readiness typically tracks technology readiness. However, deviations are possible – provided they are justified and explicitly argued for. In this way, the framework functions as a check on overtly speculative forms of ethical analysis.

5. The ethics readiness of quantum technologies

In this section, I assess the ethics readiness of quantum technologies.[8] As an emerging class of technologies, quantum technologies offer an interesting test case for the ERL framework, exploring its practical value and implications for the ethical study of emerging technologies.

Generally, quantum technologies refer to a family of technologies that leverage quantum effects to create practical applications.[9] These technologies are not entirely new; a first generation includes widely used devices like MRI scanners, lasers, and transistors. Recently, scientists and engineers have begun developing a second generation of quantum technologies, seeking to make further use of quantum phenomena by actively manipulating them in a more precise and controlled way. This promises new capabilities, such as enhanced sensing methods (*quantum sensing*), advanced computational power (*quantum computation*), and novel ways to communicate and secure information (*quantum communication*).

Although quantum ethics often speaks of 'quantum technology' as a single entity, it is important to distinguish between its subclasses, as they are at different stages of technological readiness. While not specifying exact levels here, quantum computation and communication remain at the lower end of the technological readiness scale (TRLs 1–3). Despite significant breakthroughs in fundamental research, scalable prototypes are not yet available, and their viability not yet definitively confirmed. In contrast, quantum sensing has progressed further, falling between mid (TRLs 4–6) and high readiness levels (TRLs 7–9). This technology is proven feasible and close to ready for implementation in various applications.

Drawing on the idea that technology readiness levels (TRLs) correspond to specific ethics readiness levels (ERLs), the different TRLs of quantum technologies imply that they afford different ethical inquiries. The low technology readiness of quantum computing and quantum communication indicates their generally low ethics readiness

[8] This section does not offer an ethical assessment of quantum technologies in terms of analysing their ethical aspects; rather, it assesses their ethics readiness in order to identify the most appropriate approach for such analysis.

[9] For a brief introduction to quantum technologies and the quantum phenomena they harness, see Chapter 1 (Section 2.1).

level. Given the suitability of the meaning-oriented approach to ethics at the earliest phases of development (Section 2.4 and 3.3), the ethical analysis of these technologies may be best approached through critical examination of the visions, narratives, and societal expectations currently shaping their development.

Such meaning-oriented ethics includes reflection on *whose* ideas are currently driving the development of these technology and *whom* they are intended to benefit. This brings into focus that the development of quantum computing and communication is largely driven by a small group of well-funded governments and corporations. As a result, ethical concerns arise regarding the inclusiveness of the innovation process and the eventual accessibility of these technologies (see also De Wolf, 2017). These concerns are particularly pressing at a stage when the technologies are still fundamentally in the making.

However, the question of how quantum computing and communication technologies 'come to matter' – in the sense of how they acquire meaning and significance in society – remains largely overlooked in current ethical discourse. Notable exceptions include Inglesant et al. (2021) and Possati (2025), but a broader debate has yet to emerge. Instead, ethical discussions tend to focus (prematurely) on outcomes, particularly the potential benefits and risks of these technologies, despite their early stage of development.

This tendency becomes apparent in concrete measures already being taken to mitigate specific risks – for instance, the potentially catastrophic impact of large-scale quantum computers[10] on cyber security. These efforts include exploring and preparing for a transition to quantum-safe encryption – algorithms designed to resist attacks from quantum computers and safeguard digital data (NIST, 12 November 2024; TNO, CWI, & AIVD, 2023). At first glance, this anticipatory governance may seem premature given the nascent state of quantum computing. However, the potentially severe consequences of a quantum cyber security breach may provide a rationale for anticipating specific outcomes and, in turn, for considering an outcomes-oriented ethical approach, even amid profound uncertainties (Price, 2024). Still, such an exception depends on well-supported, realistic expectations about the technology's trajectory and its potential consequences (Nordmann & Rip, 2009). In Chapter 8 I take up this thread and explore whether and under what conditions outcomes-oriented ethics may be justifiable at a low TRL.

Quantum sensing, on the other hand, occupies a higher TRL and thus a higher ERL. With its technological viability demonstrated, it affords more detailed and specific

[10] A cryptographically relevant quantum computer (CRQC) is a quantum computer with a sufficient number of reliable qubits and error-correcting capabilities to efficiently break widely used encryption methods, such as RSA or ECC, by solving the underlying mathematical problems (Mosca & Piani, 2024). See also Chapter 1, Textbox 1.

ethical analysis of impact. Outcomes-oriented ethics becomes applicable at this stage and gains relevance: demonstrated technological feasibility makes it both possible and pressing to anticipate potential consequences. However, based on the output of academic literature, the current ethical debate about quantum technologies seems to gravitate towards quantum computing and communication, while relatively neglecting sensing. This may reflect a tendency to prioritise compelling yet speculative scenarios related to quantum computing and communication, at the expense of addressing the ethical implications of the more near-term applications of quantum sensing.

The suggestion to approach the ethical discussion of various types of quantum technologies differently, illustrates how the concept of ethics readiness can guide the ethics of emerging technology. In the case of quantum, the ethics readiness framework offers a basis for redirecting the ethical discourse. By aligning the ethical approach with the stage of technological development, we can ensure that ethical discussions remain relevant and appropriately grounded at every phase of a technology's evolution.

6. Conclusion

Over the past decades, numerous methods have emerged for the ethical study of emerging technologies, each grappling in their distinct ways with the inherent uncertainties of innovation. As technologies progress, however, the nature of these uncertainties evolves, suggesting that ethical approaches should adapt to align with specific stages of technological development. Yet an overarching framework to evaluate the suitability of different ethical methods based on a technology's stage of development remained lacking.

In this chapter, I aimed to address this gap. Considering the vast number of methods that have been developed for the ethics of emerging technology, I distinguished between two main approaches: *outcomes-oriented ethics* – typically focused on the ethical assessment of consequences – and *meaning-oriented ethics* – typically focused on critically reflecting on the ideas driving technological projects as well as the broader expectations and associations attached to them. Since these approaches each deal in their own way with the uncertainty that characterise early-stage technologies, I suggested that their appropriateness and relevance depend on the availability of information, and thus on the specific stage of technological development.

To substantiate this idea, I introduced the concept of *ethics readiness*: the readiness of a technology to undergo detailed ethical scrutiny. I argued that the ethics readiness of a technology evolves along with technological maturation, determining both the effectiveness of outcomes- and meaning-oriented ethics, as well as the supported level of granularity or specificity of ethical analysis. To make ethics readiness more concrete, I defined three Ethics Readiness Levels (ERLs), in alignment with the Technology Readiness Levels (TRLs) scale. The resulting framework indicates that early-stage

technologies (TRL 1–3) typically exhibit a low ERL, reflecting the limited capacity of an immature technology to support detailed ethical analysis. At this stage, when uncertainty is fundamental and broad, the meaning-oriented approach to ethics is best suited. When a technology progresses to higher TRLs (4–9) and more information becomes available, it supports and even calls for an outcomes-oriented approach to ethics, allowing for increasingly specific ethical scrutiny.

The ERL framework captures how the appropriateness of outcomes- and meaning-oriented ethical approaches shifts as a technology matures and takes on a more concrete form. Its iterative application to the emerging field of quantum technologies served as an initial exploration of its practical implications. The analysis showed that different types of quantum technologies call for different ethical approaches, depending on their respective TRLs. This case discussion thus illustrates the framework's practical value: it helps identify the most fitting ethical approach at a given stage, and it helps recalibrate ethical engagement when this risks becoming misaligned with the technology's stage of development, for example, by raising questions prematurely.

By linking technological readiness with ethical readiness, the ERL framework ensures that ethical reflection evolves in tandem with technological advancement and the understanding available of what the technology can do and may be used for. It offers a dynamic, pragmatic approach to upstream ethics, aligning ethical reflection with the affordances of technology's maturity. In doing so, it resolves the tension between engaging too soon and too late, and helps navigating between premature speculation and missed ethical opportunities. That way, the concept of ethics readiness lays the groundwork for a more responsive and impactful upstream ethics.

Finally, this chapter opens up avenues for further discussion on at least two fronts. The first concerns the correlation between a technology's maturity and its ethics readiness. This correlation should be understood as a heuristic: a guiding principle rather than a strict rule. Deviations are possible, provided they are well justified. This chapter suggested that one potential justification is the severity of the technology's potential impact: severe enough consequences may warrant engaging with outcomes even at an early stage. Chapter 8 takes up that thread and develops it into a more specific question: under what conditions is it ethically defensible to treat a still-speculative technology as posing a risk that warrants precautionary action? Second, in advocating a meaning-oriented approach to the ethics of quantum computing and communication, this chapter points to the importance of scrutinising the sociopolitical dimensions of these technologies, in particular, the ideas and expectations that shape their development. This is the focus of the next chapter, which examines the visions of key actors in the quantum computing ecosystem.

Chapter 6

Visions of Quantum Computing

An earlier version of this chapter was published as: de Jong, E., & Grunwald, A., (2026). Visions of Quantum Computing: An Ethical Assessment. Futures, 103941. DOI: https://doi.org/10.1016/j.futures.2026.103941

Abstract

In the previous chapter I argued that the ethical analysis for early-stage technologies like quantum computing is best approached through studying the ideas driving their development. Visions are a specific type of such ideas. They are normative ideas about future technologies, shaping their development by mobilising resources and influencing decisions. This productive aspect of visions makes them salient objects of ethical analysis. This chapter therefore examines contemporary visions of quantum computing and the futures they project. First, I analyse vision statements by key actors in the field, focusing on anticipated technical affordances, the applications they enable, their associated domains, and the value-laden terms used to describe these projections. Drawing on these insights, I distil four main visions of quantum computing, each framing its societal relevance differently: as a contributor to the public good, a scientific tool, an economic asset, or a geopolitical advantage. Second, I highlight ethically salient patterns and presuppositions embedded within these visions. These include a pervasive techno-optimism paired with technosolutionist reasoning; the prevalence of promissory language used to legitimise innovation agendas; and a morally ambivalent discourse shaped by the interests of a narrow group of powerful actors. Taken together, these features form a figurative heatmap of potential ethical concerns in contemporary visions of quantum computing that warrant closer ethical scrutiny and point to avenues for further research.

1. Introduction

Emerging technologies often come with grand promises. These promises are wildest in the earliest phases: when their full potential is still unknown, technologies are frequently framed as promising 'transformative forces' that could reshape society – for better or worse. These early visions are expansive because the technology itself is still open-ended. When everything is still possible, well, everything is still possible. Quantum technologies exemplify this dynamic.

While first-generation quantum technologies – such as MRI, lasers, and transistors – are already embedded in society, a second generation is now being developed. Among these innovations, quantum computing has emerged as the principal focus of attention (for a brief introduction to quantum computing, see Chapter 1, Section 2.1). The expected capabilities of large-scale quantum computers have generated high hopes for applications across diverse domains, from accelerating drug discovery and improving climate modelling, to revolutionising artificial intelligence and challenging existing cyber security systems.

High expectations bring quantum computing into the realm of ethics and responsible innovation[1]: How can this technology be guided towards societally desirable outcomes? How can we ensure "the right impact"? (Von Schomberg, 2013). Since Vermaas' special issue on the societal impact of emerging quantum technologies (2017), an ethical discourse around quantum computing has begun to take shape. A growing body of literature highlights the potential societal consequences of this technology and emphasises the need to guide its development responsibly (De Wolf, 2017; Inglesant et al., 2021; de Jong, 2022; Ten Holter et al., 2023; Seskir et al., 2023; Possati, 2023; Kop et al., 2024a, 2024b; Gasser, de Jong, & Kop, 2024).

Notwithstanding abundant ambitions, high expectations, and efforts to anticipate its societal impact, a practically relevant quantum computer – capable of solving real-world problems more effectively than classical alternatives – remains elusive at present. Paradoxically, what is currently technically possible holds little societal significance, while what is societally significant remains technically out of reach.[2] At present, quantum computing exists primarily as a projection: a (scientifically grounded) technological promise.

Building on the notion of ethics readiness (Chapter 5), this promise-laden status of quantum computing has implications for its ethical analysis. Much of the current ethical discourse centres on assessing quantum computing's anticipated impacts,

[1] Here understood as anticipating the ethical, legal, and social implications of technological innovation.

[2] For example: Google's demonstration of quantum advantage, while a milestone in computational capability, has no immediate practical applications; while cryptographically relevant quantum computers – those powerful enough to break existing encryption or enable other transformative applications – would require a vastly greater number of stable qubits than any system currently available.

considering the benefits and risks that might arise once the technology is realised. However, this tends to involve engaging with highly hypothetical futures, speculating about what the technology *might* become and what it *might* do. Such "speculative ethics" has been widely criticised for potentially diverting intellectual resources towards a mirage (Nordmann, 2007; Nordmann & Rip, 2009; Grunwald, 2010).

As the previous chapter argued, to avoid speculative overreach, ethical enquiry into early-stage technologies like quantum computing should shift the focus from the future technology itself, to *how it is being envisioned* (Chapter 5): How is the societal relevance of quantum computing imagined? By whom? And under what assumptions? These are ethically salient questions that this chapter engages with. Long before a technology materialises, the meanings projected onto it – through ambitions, expectations, hopes, and fears – can have tangible effects in the present (Michael, 2000; Borup et al. 2006; Pyrrho, 2011; Ferrari, Coenen, & Grunwald, 2012; Dignum, 2018; Schneider et al., 2023). In other words, prior to analysing technological artefacts themselves, we can examine the ways in which they are imagined and prospectively endowed with meaning (Wynne, 2007; Ferrari, 2013).

Among such anticipatory imaginaries, *visions* play a particularly important role. Technological visions are active, strategic narratives that frame the future of a technology and define its supposed societal relevance (Grunwald, 2018; Schneider et al., 2021). Visions are never neutral: they are inherently normative because they outline futures considered worth pursuing or avoiding[3]. In articulating such desirable and undesirable futures, visions give the technologies involved *meaning*. In other words, visions 'make sense' of technologies in specific ways (Kudina, 2021), assigning them roles and significance in relation to what is valued in those futures. This value-ladenness, in combination with their capacity to shape technological trajectories, makes visions important objects of ethical analysis.

Despite this significance, visions of quantum computing have received limited ethical scrutiny (Alleblas, 2026, p. 6). Notable exceptions include Roberson (2021) and Possati (2024), who examine societal meanings in projections of quantum technologies more broadly, and Inglesant et al. (2021), who explore expectations around technical capabilities and timelines. Yet a systematic study of the societal meanings specifically attributed to quantum computing, and an ethical assessment of those meaning-attributions, remains lacking.

This gap, together with the conclusions of the previous chapter, provides the rationale for the present study, asking: In what ways is quantum computing envisioned? How is

[3] This understanding of visions also accommodates 'dissident visions', i.e. imaginaries of undesirable futures and the pathways proposed to prevent them. Chapter 8 will discuss this statement in more detail, casting the risk associated with large-scale quantum computing as a special type of sociotechnological vision.

its societal significance constructed, and what are the underlying presuppositions? To address these questions, I critically examine contemporary visions of quantum computing and highlight the ethical questions they raise at this formative stage. After outlining my approach of *ethical vision assessment* (Section 2), I undertake a two-stage analysis. First, I analyse vision statements from across key actors in the field and distil from them four main visions of quantum computing and its societal relevance (Section 3). Second, I analyse these visions from an ethical perspective, highlighting patterns and presuppositions embedded within these visions that (can) give rise to ethical concerns (Section 4). This produces a figurative heatmap of ethically salient issues that warrant closer scrutiny.

The chapter thus offers an empirical analysis of how the future of quantum computing is currently being envisioned by those driving its development, and an ethical analysis of these visions, revealing continuities with broader patterns and issues in the ethics of new and emerging science and technology (NEST-ethics). The resulting ethical heatmap sets an agenda for further research.

2. Method: Ethical Vision Assessment

To scrutinise contemporary visions of quantum computing, this study adopts – and adapts – the approach of *vision assessment*. Originally introduced by Grunwald (2018), this evolving approach aims to uncover the underlying presuppositions and normative commitments embedded in technological visions. As a hermeneutic method, it examines how technologies acquire meaning by analysing the visions through which that meaning is constructed (Grunwald, 2020; Nordmann & Grunwald, 2023).

In this chapter, I further develop and extend vision assessment by making its ethical dimension explicit, focusing on the ethical issues raised by emerging visions of quantum computing. To operationalise this *ethical vision assessment*, I adopt a two-stage analytical strategy, combining elements of reflexive thematic analysis, discourse analysis, and critical hermeneutics. First, I map contemporary visions through a thematic analysis of a corpus of 'vision statements'. Second, I interrogate their ethical dimensions: Who is shaping these visions? Whose interests do they serve? What values do they promote? And what assumptions do they make about what is desirable, inevitable, or urgent in the imagined future of quantum computing? I elaborate on each of these stages below.

2.1 Stage one: identifying visions

2.1.1 Materials selection

To identify and analyse contemporary visions of quantum computing and its societal significance, publicly available texts were examined – such as reports, policy

documents, and other web-based publications – that articulate expectations about the future of quantum computing and, taken together, reflect the contemporary discourse.

The selection of texts was guided by two criteria. First, I deliberately focused on actor groups that are currently and actively shaping the development of quantum computing and its surrounding discourse. Five such groups were distinguished: governments, coordinating organisations, companies, research institutes, and thought leaders (prominent individuals). Within each group, specific actors were selected as follows:

- *Governments:* Countries and supranational unions with published national or regional quantum strategies.
- *Industry:* The five largest tech firms active in quantum computing and five prominent companies specialising exclusively in quantum technologies.
- *Coordinating organisations:* Organisations that do not develop quantum computing technologies themselves but play a key role in shaping strategic direction, aligning actors, and influencing innovation and policy landscapes.
- *Research institutes:* Research centres with an explicit focus on quantum computing.
- *Thought leaders:* Individual researchers widely recognised as prominent voices in international debates about the future of quantum computing.[4]

Looking through a lens of the quadruple helix, an expectable additional actor group would be 'society', as represented, for example, by civil society or the broader public. I deliberately choose to focus on the five actor groups above, for two interrelated reasons. The first is their influence. The visions of these actors do not merely express desired futures, but also set agendas and mobilise resources to pursue these futures. This makes their visions particularly ethically salient, not because they are exhaustive, but because they are influential in setting agendas and expectations.

I recognise that focusing on such "vanguard visions" promoted by elite actors has been criticised for sidelining alternative voices (Sand, 2019). At present, however, these alternative voices have not yet fully developed within the discourse on quantum computing. That is the second reason for focusing on the current groups. Quantum computing remains a very early-stage technology that has not yet convincingly entered public debate or civil-society agendas. Empirical studies indicate that public discussion of the societal impacts of quantum computing remains limited (Van de Merbel, 2024; Meinsma et al., 2025). References to quantum computing in (news) media often report on scientific developments rather than reflecting an emerging public discussion. While a small number of scholars have adopted more critical perspectives, they do not yet form a coherent actor group shaping the technology's development. Where appropriate,

[4] This selection was based on my experience in the field.

I engage with such critical work in our ethical analysis of the identified visions. I thus see the focus on vanguard visions justified not just by their power, but also by the current empirical absence of broader societal imaginaries. I further reflect on this absence in Section 4.

The second criterion for the selection of the corpus was to select for each actor textual material that most elaborately articulated claims about what quantum computing[5] is expected to be, what it will do, and why this matters. I refer to these texts as the actor's *vision statements*: public statements about the imagined future of quantum computing and its meaning for society[6].

The corpus was assembled through an extensive desktop search conducted in May 2025[7] and resulted in a corpus of 64 actors and their vision statements (see Appendix). It includes national strategies, action plans, roadmaps, agendas and initiatives in the case of governments; reports, papers, and web-based publications in the case of coordinating organisations; website texts, mission statements, and organisation descriptions in the case of companies and research institutes; and blogs and interviews in the case of thought leaders.[8] I treat this corpus as reflecting the current discourse among prominent actors about the significance of quantum computing and its meaning for society.

2.1.2 Thematic analysis

In the first analytical step, I conducted a qualitative content analysis aimed at identifying recurring themes across the vision statements. I chose not to study the visions per actor group, but instead to focus on patterns across the corpus. This choice was motivated by the aim of this study, which is to map the discursive landscape shaping quantum computing's trajectory as a whole, and not to evaluate or compare specific actors' positions. For that purpose, what matters most is which framings and issues are prevalent across the discourse; an actor-by-actor breakdown would answer a different question than the one the current analysis is designed to address. Additionally, I wished to avoid essentialising actor groups – treating categories such as 'government' or 'industry' as internally coherent and externally distinct – when in practice there is likely to be considerable heterogeneity within, and overlap between, these categories.

[5] In this analysis, I included 'quantum simulation' in the understanding of 'quantum computing'.

[6] I take 'society' as an overarching term, including also the economy and the environment.

[7] The corpus necessarily reflects a time-bound snapshot of discourse.

[8] I recognise the difficulty of reliably attributing documents to a single actor or actor group: strategic documents, for instance, are often produced collaboratively and may reflect input from multiple types of actors. This does not pose a problem for the analysis, however, since I focus on patterns across the discourse rather than on specific actors or actor groups. Where I refer to documents as associated with a particular actor group, this denotes the document's institutional origin rather than a claim about its individual authorship.

This also would require a different methodological set-up. A corpus-level analysis avoids this kind of overattribution, while still allowing us to identify which framings and issues are prevalent within the discourse as a whole.

The analysis was inspired by the approach of thematic analysis (Braun & Clarke, 2006). After an initial close reading of the selected materials, I selected text excerpts that articulated ideas about quantum computing and the future it enables. Depending on the document, this resulted in the selection of multiple paragraphs or, in some cases, only a few sentences.

From this initial reading, I inductively defined four thematic categories that naturally structure the material: (1) the perceived *technical affordances* of quantum computing; (2) its practical *applications*; (3) the *domain* of application; and (4) the *value-oriented language* used to frame these possibilities:

1) *Technical affordances:* The perceived capabilities or functions of quantum computing (e.g. "modelling complex molecules", "enhanced computational abilities", "optimisation"). This category helps isolate the technical narrative from broader societal framing of the technology.
2) *Application:* This captures the immediate practical goal associated with the technology, such as "designing novel batteries" or "improving encryption methods". It serves as a bridge between technical capabilities and societal domains.
3) *Domain:* The sector or societal field in which the applications are situated, linking the technology to particular 'problem spaces', like "health", "energy", or "defence". This provides context for how actors imagine the societal significance of quantum computing.
4) *Value-oriented language:* Evaluative and normative framings through which quantum computing is presented, e.g. as desirable, necessary, risky, or consequential. Such language, implicitly or explicitly appealing to values, expresses judgments about what ought to be pursued, prioritised, or avoided. In some sources, value-oriented language is explicit (e.g. "sustainable", "improve", "revolutionise"), directly framing quantum computing as serving social or ethical aims. In other cases, explicit value language is absent, but underlying values can still be inferred from the way applications or societal impacts are framed. For example, references to using quantum computing for better battery design may implicitly invoke the value of sustainability, given its association with clean energy. I therefore coded value-oriented language using the subcodes 'explicit' and 'implicit'.

These categories or themes are analytically useful as they help distinguish between what quantum computing *is imagined to be*; what it *is imagined to do*; and what it *is*

imagined to do good for. For each selected excerpt, I highlighted substantive phrases, and coded these across the four themes.

Initial qualitative analysis of the coded material revealed that quantum computing was almost universally envisioned as bringing about some kind of change. As I deemed this characterisation meaningful for a deeper understanding of visions of quantum computing, I subsequently added *change* as an additional code.

Table 1 gives an overview of the complete coding scheme, illustrated by examples.

Section 3.1 presents the analytical results per theme. By qualitatively analysing and interpreting these results, I distilled four main contemporary visions of quantum computing and its societal relevance, which I discuss in Section 3.2. These visions reflect interpretative choices made in this study, particularly the choice to focus on quantum computing as a driver of change and to look at projections *across* different actor groups rather than per actor category (see also footnote 8). I thus do not claim that these four visions are the only possible ones; different analytical choices may plausibly produce a different mapping.

Table 1. Overview of themes (codes), their description and illustrative examples from the corpus (table continues on the following pages)

Theme (code)	*Description*	*Example*
Technical affordances	Claims about the anticipated technological capabilities of quantum computing	"We expect that [quantum computers] will be able to solve some problems exponentially faster than classical computers can." (25). "It promises problem-solving capabilities far beyond classical computing, particularly for complex simulations and optimisation." (37). "The three types of problems where quantum computing will have a transformational impact are: the simulation and modelling of quantum systems and processes occurring in nature; search for the ideal or 'optimal' solution to a given problem with multiple existing answers; and algebraic problems such as factorization." (41).
Applications	Claims about how quantum computing is expected to be applied	"Material design (e.g. new battery materials), pharmaceutical development, climate change modelling and environmental monitoring may be areas where progress will be made." (8) "Applications include the development of new sources of clean energy, monitoring climate change on the planet, developing new vaccines more quickly and accurately, and secure data communication." (3). "The anticipated applications for this technology range from drug discovery, new materials and new ways to tackle climate change to financial

Table 1. (continued)

Theme (code)	*Description*	*Example*
Applications *(continued)*	Claims about how quantum computing is expected to be applied *(continued)*	services, cybersecurity and a quantum internet." (22). "Some applications where quantum computers can provide a speed boost include machine learning (ML), optimization, and simulation of physical systems." (28). "Quantum computing promises to enable breakthroughs in materials, chemicals and drug design, financial and climate modelling, and cryptography." (29). "For many scientists, the most exciting quantum computing application is the opportunity to study quantum physics itself in a way that wasn't possible before." (50). "Quantum computers are expected to challenge current cryptography methods and to introduce new possibilities for completely private communication." (58).
Domain	Claims about the domains in which anticipated applications are situated	"Quantum technology holds enormous potential for science, the economy, and society." (11). "The development of quantum computers [can] create opportunities for all sectors of the economy." (15). "We foresee a significant presence of end-users from sectors like agriculture, energy, security, science, and digital infrastructure." (38). "The long-term goal is to develop fault-tolerant quantum computers which [...] are of broad use for both the economy and society." (43). "Quantum technology is [...] considered a potential game-changer in many social and economic sectors, including health, agriculture, climate, and safety." (56).
Value-oriented language	Evaluative and normative terms or phrases that qualitatively appraise the anticipated impacts of quantum computing	*See below*
Explicit	Explicitly stated qualitative appraisals (e.g. clear reference to values or normativity)	"[T]his transformation will: modernise our economy, improve our society, support national interests." (1). "[We must] recognise the strategic importance of quantum technologies for the scientific and industrial competitiveness of the EU." (7). "Quantum technologies have the potential to benefit society and almost all industries." (18). "To accelerate scientific discoveries for a better future." (27). "[Our mission is] to accelerate quantum computing and use its power to positively transform the world." (33).

Table 1. (continued)

Theme (code)	*Description*	*Example*
Explicit *(continued)*	Explicitly stated qualitative appraisals (e.g. clear reference to values or normativity) *(continued)*	"[Quantum computing] promises to bring transformative benefits to several industries and have a considerable impact on improving the state of the world in the coming years." (41). "Quantum computer as a research tool […] whose results benefit humankind." (44).
Implicit	Implicit or indirectly expressed qualitative appraisals (e.g. adjectives or framings)	"The future is quantum." (44). "Quantum is going to change the world." (34). "Now we have another opportunity to be among the world's leading forces in a new technological revolution." (8). "Once we master quantum technology, it will transform our world." (15). "We need way more [computational] power." (30). "We believe that quantum technologies can be a game changer in many societal and economic sectors, including health, agriculture, energy and safety." (45). "Although we don't know when we will build a quantum computer that will threaten current public-key ciphers, it will happen one day." (64).
Change	Characterisation of quantum computing as a force of change	"The paradigm-shifting computational differences of quantum computing give rise to extraordinary transformational opportunities." (41). "Quantum technologies will make a difference improving economic and national wellbeing." (1). "This emerging field will transform how we develop and design everything from life-saving drugs to next-generation batteries." (4). "With new tools and new ways of thinking, quantum computing will forever change the way we solve problems across industries." (31).

2.2 Stage two: ethical analysis

In the second step of the analysis, I subject the identified visions to an ethical analysis or 'assessment'. I adopt a critical hermeneutic approach (Timmermans & Blok, 2021), aiming to develop an understanding of the meaning assigned to quantum computing and to uncover recurring patterns, presuppositions and assumptions embedded in these visions. Drawing on tools from hermeneutic analysis of visions (Grunwald, 2018; Sand, 2023), the analysis is structured across three analytic levels:

1) *Content (substantive level):* This level focuses on presuppositions and assumptions about desirable futures and the role assigned to quantum computing in attaining them. The central ethical question is: *Why is quantum computing mobilised in these visions?*
2) *Framing (discursive level):* This level examines how visions are rhetorically constructed and morally legitimised. The central ethical question is: *How are these visions framed as normatively compelling?*
3) *Author (agential level):* This level considers who produces and promotes these visions and which interests or stakes may underlie them. The central ethical question is: *Who advances these visions, and why?*

Using these levels as analytical lenses, I identify issues that constitute, or may develop into, ethical concerns warranting further scrutiny. The results of this analysis are presented in Section 4 ("Ethical assessment of quantum computing visions"). First, I discuss the visions of quantum computing in the next section.

3. Visions of quantum computing

This section presents the results of the first stage of our analysis. Section 3.1 discusses the findings for each analytical dimension – value-oriented language, domains, applications, and technical affordances – and highlights some meta-level observations. Section 3.2 then synthesises these findings and identifies four dominant visions of quantum computing.

3.1 Findings: value-oriented language, domains, applications, and technical affordances

3.1.1 Value-oriented language: a high-potential technology enabling a desirable future

A key characteristic of value-related language surrounding quantum computing that I identified is its strong *orientation towards the future*. Across the material, quantum technologies are frequently associated with promissory terms such as "potential", and expressions like "unlocking" or "realising potential". These formulations present the value of quantum computing as largely deferred, that is, not located in present capacities, but expected to emerge through continued technological development. The following quotes are illustrative:

> "Quantum technology is an exciting and rapidly evolving field of scientific research and innovation with tremendous disruptive potential in the medium to long term. […] Quantum computing may unlock solutions and new responses to specific challenges across

> many sectors and in turn may accelerate progress towards several of the United Nations Sustainable Development Goals." (22)
>
> "[Denmark has] a unique opportunity to unlock the immense potential of quantum technologies, not only in the realm of research but also in translating this knowledge into solutions that can benefit society, businesses, and national security, ultimately improving the lives of individuals worldwide." (6)

Secondly, this anticipated value is often described in terms of *significant change*. Common terms include "transformation", "revolutionise", "game-changer", "paradigm shift", and "disruption", even "changing people's everyday lives", framing quantum computing as a "frontier technology" that does not merely contribute to existing trajectories but rather alters or reshapes them. The areas associated with such change include the economy, science, national security, geopolitics, and society at large:

> "Quantum computing has the potential to revolutionize the way we solve problems that are impossible to solve using classical computers. It is therefore expected to have a significant impact on a great number of fields, from cybersecurity and chemistry to finance and climate science." (9)
>
> "Quantum computing is recognized as one of the most critical breakthrough technologies, with the potential to radically alter society." (23)
>
> "By applying the laws of quantum physics to computing, we will achieve unprecedented breakthroughs in drug discovery, healthcare, materials science, cybersecurity, energy transformation, and more." (33)
>
> "The paradigm-shifting computational differences of quantum computing give rise to extraordinary transformational opportunities." (42)

Third, the projected change is largely cast in terms of positive development, indicated by the ubiquity of terms like "advancing", "enhancing", "accelerating", "improving", "benefitting", and "breakthrough". In this way, quantum computing is frequently framed as an "opportunity", an "advantage", an "enabler" or "key" to reaching desirable goals across multiple sectors:

> "If we are successful, quantum innovation could be a positive disruptive power that leads to improved lives, a richer, more sustainable economy, a healthier planet – and an even deeper understanding of our universe." (38)
>
> "Society and almost all industries [...] stand to benefit from quantum technologies due to the exponential increase in computational power and secure communication capabilities." (18)
>
> "Quantum computing [...] promises to bring transformative benefits to several industries and have a considerable impact on improving the state of the world in the coming years." (41)

> "Quantum technologies already offer possible solutions to some of our greatest societal challenges and, perhaps most excitingly, offer future capabilities that are yet to be explored. They will improve lives, drive economic growth and create jobs, and make us more secure." (17)

A notable exception to the mostly positive projections of a quantum-enabled future is the flagging of the potential "risks" of quantum computing for security, in particular its "threat" to currently widely used encryption schemes that keep digital data secure:

> "Quantum computing [has] the potential to radically alter society - affecting [...] the nature of threats to national and personal security, and governance systems."(23)

> "Quantum technologies have a double-edged impact on cyber security and defence, benefitting both the defensive as well as the offensive side." (42)

> "Quantum technologies will very likely change our approach to digital security. [...] So current encryption methods need to be changed." (64)

Fourth, references to progress are often linked to a *wide range of value domains and societal goals*. These include security, health, sustainability, economic growth, scientific advancement, curiosity, efficiency, and social wellbeing. The desirable futures associated with quantum computing are thus 'multi-driven', incorporating both instrumental goals (e.g. efficiency, competitiveness) and broader societal aims (e.g. sustainability, wellbeing):

> "Quantum computing can help develop solutions to major societal challenges, such as health care, energy production and climate change research." (8)

> "The broad field of quantum science and technology holds tremendous potential for science, the economy and society." (12)

> "Quantum computers are a revolutionizing technology – they have the potential to transform business, society, and the planet for the better." (32)

> "The first companies to use quantum computing to realize improvements in cost, speed, or accuracy [...] could see significant changes to their bottom line, save years of development time bringing new products to market, or break ground in new industries yet to be imagined." (31)

Finally, and closely related, quantum computing is often framed in terms of strategic importance. It is portrayed as a key asset in national competitiveness and international leadership, with references to a "global race", "technological sovereignty", and efforts to "position" countries or regions at the forefront of scientific and technological power. This framing adds a sense of urgency to the broader value discourse, reinforcing the perception that securing quantum capabilities is vital for maintaining strategic advantage and addressing national interests. This is illustrated by the following quotes:

> "Quantum technology is one of Finland's future industries. We have every opportunity to be global leaders in this area." (8)

> "It is therefore obvious that quantum technology is closely tied to immense momentum in the global innovation race. [...] How we deal with the topic of quantum computing is indeed also a question of technological sovereignty in global competition with strong rivals." (11)

> "To remain competitive on the global stage, Lithuania needs to actively engage in the development and application of QT." (23)

3.1.2 Domains: everything, everywhere, all at once

The analysis also revealed that, at a high level, quantum computing is projected as a technology of wide societal relevance, with anticipated impact across almost every major domain in society, and "a wide range of application fields" spanning the entire economy. Based on recurring themes and framings, four main domains can be distinguished where quantum computing is imagined playing a significant role: society, the economy, science, and national security.

Society – First, quantum computing is frequently envisioned as a transformative tool for addressing societal challenges and achieving shared goals. It is associated with a wide range of application areas, including healthcare, energy systems, environmental sustainability, agriculture, cyber security, and defence. Frequently, its potential is explicitly connected to the United Nations Sustainable Development Goals (UN SDGs), with the technology framed as a contributor – sometimes even a key enabler – of progress on global health, clean energy, and environmental protection. This societal framing tends to emphasise both public benefit and the pursuit of long-term sustainability. The following quote is illustrative here:

> "[Our mission is] to promote global and inclusive access to quantum computing, developing applications that benefit humanity and support the United Nations Sustainable Development Goals (SDGs)." (36)

Economy – Second, in the economic domain, quantum computing is commonly positioned as a driver of national and industrial competitiveness. It is linked to job creation – particularly in high-skill sectors – economic growth, and innovation across diverse industries such as logistics, manufacturing, AI, and finance. This includes both enhancing existing capabilities and enabling new ones (see Sections 3.1.3 and 3.1.4), with quantum computing often described as a foundation for a future digital or "quantum economy". The following quote illustrates this:

> "Quantum technologies will make a difference improving economic and national wellbeing. [...] Commercialising quantum technologies could create an Australian quantum industry worth $2.2 billion and directly employing 8,700 people by 2030. This could reach approximately $6 billion and 19,400 jobs by 2045. These technologies are expected to add billions of dollars of value to related industries in that timeframe." (1)

Science – Third, quantum computing is expected to advance scientific discovery. It is projected to open up new frontiers in research areas such as chemistry, physics, materials science, and biology. In this projection, quantum computing is imagined as a tool for overcoming limitations of research tools based on classical physics, enabling breakthroughs that are currently out of reach:

> "Our mission [is] to accelerate scientific discoveries for a better future." (27)

> "Well, remember I'm a scientist, so I like to think of computers as tools for advancing science, for scientific discovery. We use computers to illuminate how nature works in a variety of ways. […] Quantum computers will be good for solving problems like that." (6)

National security – Fourth, national security emerges as a distinct and prominent domain where quantum computing is expected to have significant impact. At one level, it introduces a strategic and geopolitical dimension to the technology's anticipated economic value, with quantum computing portrayed as a critical asset in securing technological sovereignty and positioning within global competition. At another level, it is expected to influence (inter)national security more directly – particularly through applications in defence and cyber security – where it is viewed both as an opportunity (e.g. enabling advanced encryption and secure communication) and as a threat (e.g. the potential to undermine existing cryptographic systems):

> "Quantum computers […] could also have far-reaching implications for national and international security, as they enable new defence capabilities and threaten current data protection systems." (35)

3.1.3 Applications: quantum (r)evolution

Before detailing specific applications, one overarching pattern that clearly emerges in how quantum computing is envisioned is the tension between *evolution* and *revolution*. On the one hand, applications are framed as the next step in a continuum of technological progress, representing incremental enhancements of existing processes. On the other hand, quantum computing is also positioned as a fundamentally disruptive technology, imagined unlocking capabilities beyond the limits of classical computing, sometimes described in almost aspirational terms as "making the impossible possible".

Zooming in on specific areas of envisioned application, several clusters emerge as the most frequently mentioned and substantively developed:

Healthcare and life sciences – Quantum computing is anticipated to revolutionise drug discovery, accelerate medical research, and improve diagnostics and personalised treatments. Envisioned applications include molecular simulation, protein folding analysis, pharmacokinetic modelling, and advanced medical imaging. Quantum computing is frequently framed as a key enabler of faster, more targeted healthcare innovation:

> "The exponential increase in computing power from quantum computers could revolutionise our healthcare system - from dramatically improved drug discovery techniques to providing personalised treatment to an individual based on genetic and environmental factors." (17)

Environmental and energy solutions – Quantum computing is expected to support climate and environmental goals through improved modelling, monitoring, and optimisation. Key applications include energy-efficient systems, clean battery and catalyst design, smart grid management, emissions reduction, reduction of 'forever chemicals', precision agriculture by enhanced weather forecasting and crop management, and the development of sustainable fertilizers and pesticides. This quote is illustrative:

> "Anticipated benefits [include] improved weather and climate modelling, pollution remediation, climate-friendly industrial processes." (37)

Finance and business analytics – In financial services, quantum computing is expected to be useful ("enhance") risk management, fraud detection, scenario planning, and portfolio optimisation. It is also linked to advances in AI acceleration, big data analytics, and transaction security, envisioned to improve efficiency and resilience in business operations and financial markets:

> "With new tools and new ways of thinking, quantum computing will forever change the way we solve problems across industries. [For example,] optimizing returns and risks for large financial portfolios." (31)

Logistics, mobility, and infrastructure optimisation – Quantum computing is envisioned as a solution for complex optimisation problems in transport, logistics, and infrastructure planning. Proposed applications include traffic flow management, route optimisation for freight and emergency services, smart city design, and improved supply chain efficiency and resilience across sectors:

> "By harnessing these capabilities, quantum computers could [...] optimise logistics to deliver our parcels faster and more accurately." (1)

Manufacturing and materials development – In manufacturing, quantum computing is expected to optimise production processes and enable the design of novel materials. Applications include the development of new batteries, catalysts, and corrosion-resistant components. Quantum computing is often positioned as a driver of more efficient, sustainable industrial innovation:

> "By enabling the simulation of new materials and chemical reactions, quantum computing can contribute to the creation of better batteries and more efficient energy systems." (15)

Scientific discovery – Quantum computing is framed as a transformative research tool in fundamental science. It is projected to accelerate discoveries in physics, chemistry, biology, and materials science through powerful simulation capabilities, opening up

new avenues for exploring quantum systems, molecular interactions, and other theoretical problems previously beyond reach:

> "Quantum computers will help us learn about, model, and manipulate other quantum systems. That ability will improve our understanding of physics." (58)

National security and cyber security – In security domains, quantum computing is depicted as both a strategic asset and a disruptive threat. Envisioned applications include quantum-enhanced encryption, secure communications, cyber operations, and military planning, as well as concerns about breaking existing cryptographic systems and undermining digital infrastructures:

> "If fully adopted, functional quantum technologies would allow private and public actors in the Alliance to better protect their data and communications in a way that is fast and reliable. […] A functional quantum computer would also have the ability to break current cryptographic protocols." (42)

Risk, forecasting, and emergency management – A cross-cutting cluster highlights quantum computing's role in anticipating and responding to systemic risks. Envisioned applications include climate and weather forecasting, flood and seismic risk assessment, emergency response modelling, and food security prediction, suggesting potential for enhanced preparedness and resilience in crisis management. The following quote is illustrative:

> "Applications of quantum computing explored at OQI focus on […] food security - last mile delivery […], resilient global food supply […], weather and climate forecasting." (36)

These envisioned applications hinge on the specific technical capabilities that are attributed to quantum computing, which is discussed next.

3.1.4 Technical affordances: transcending classical capacities

The analysis reveals that the anticipated technical capabilities – or *affordances* – of quantum computing are commonly framed around four central functions: *simulation, optimisation, modelling,* and *solving computationally intractable problems*. These affordances are often attributed to the inherent affinity between quantum computers and natural quantum systems, as well as to their anticipated ability to exploit quantum-mechanical properties to solve certain classes of computational problems – from large-scale search to structural problems such as factorisation – more efficiently than classical systems. This is illustrated by the following quotes:

> "Quantum computational technologies present the opportunity to tackle many practical problems that are currently infeasible to solve, even with the most powerful classical supercomputers." (41)

> "Because quantum computers use quantum physics to access new computational abilities that are inaccessible to classical computers, we expect that they will be able to solve some problems exponentially faster than classical computers can." (26)

These perceived capabilities are often framed as *quantum-specific,* which underpins the portrayal of quantum computing as a "key enabling technology" capable of transcending classical limitations and providing solutions to problems previously deemed unsolvable. Quantum simulation, in particular, is frequently cited as a distinctive capacity, whether for modelling molecular interactions, physical systems, or complex environmental processes.

When it comes to the feasibility of these affordances, some 'vision authors' are more cautious or modest than others. The majority portrays quantum computing as a paradigm-shifting force, capable of exponentially accelerating problem-solving and enabling 'entirely new possibilities':

> "Quantum research opens up entirely new possibilities that will eclipse anything previously possible. Major leaps in innovation are on the horizon." (2)

On the other hand, more measured assessments reveal the speculative nature of these claims. They note the absence of a clear empirical basis for some of the stronger expectations, the lack of practical demonstrations of quantum advantage in most real-world applications, and the risk of overhyping what remains a nascent and highly experimental field:

> "It's important to understand that it's still an open question where quantum computers will find their niche. Potential applications to drug development and logistics optimization are possible and understandably exciting, but we're still in the early days of this computing revolution. [...] Scientists have come up with a handful of killer apps, but for the most part no one yet has a clear picture of the full capabilities of quantum computers." (50)

Indeed, while quantum computing is often presented as inherently "superior to traditional methods", this superiority is frequently restricted to a narrow class of problems for which specific quantum algorithms may yield a speed-up, and whose practical significance is still under debate.

3.2 Quantum computing and the promise of change

While the framing of quantum computing as a driver of change was a clearly recurring pattern, differences were notable in the way this change is framed. This observation is reminiscent of the familiar phenomenon of 'ambivalence' that has often been noted in relation to promising technologies (Swierstra & Rip, 2007; Shelley-Egan, 2010): the coexistence of competing framings of the same technology. Applied to change, this means that the change a technology is expected to bring may be framed in different –

sometimes contradictory – ways, for example as "revolutionary" or as "business as usual" (Swierstra & Rip, 2007, p. 9). I inductively distinguished three modes of change associated with quantum computing. Rather than standing in direct opposition to one another, these modes reflect different framings of the same technology.

The first is *revolutionary change*, signalled by terms such as "disruption", "transformation", and "game-changer". In this mode, quantum-induced change is characterised as fundamentally transforming existing systems and fields such as security and materials science:

> "While quantum computing is still in its infancy, it has the potential to revolutionize industries such as finance, healthcare, and logistics." (9)
>
> "Quantum technologies have potentially revolutionary and disruptive implications, which can degrade the Alliance's ability to deter and defend." (42)
>
> "We are creating the quantum future, because we believe that quantum technologies can be a game changer in many societal and economic sectors." (45)
>
> "Quantum technologies are not just an incremental step. They are a transformational step – a revolution that will impact our world." (46)

Second, I distinguish *augmentative change*, signalled by terms such as "enhancing" and "improving". Here, the change associated with the arrival of quantum computers is characterised as scaling existing capabilities or making operations more efficient:

> "In machine learning, quantum computing enhances the processing of large training materials." (8)
>
> "Quantum computing has the potential to revolutionise healthcare and medical science [through] accelerating drug discovery." (23)
>
> "Quantum computing could help make hospital systems more efficient, improving delivery and reducing costs." (17)
>
> "Making research in chemistry and materials science faster, easier, and more productive." (27)
>
> "Quantum computers will improve the accuracy of long-term climate models." (39)

A third mode of change is catalytic, signalled by terms such as "enabling" and "unlocking", and portrays quantum computing as enabling previously infeasible possibilities:

> "Quantum technology offers a promise to make the impossible possible." (8)
>
> "Quantum research opens up entirely new possibilities that will eclipse anything previously possible." (2)
>
> "Quantum computational technologies present the opportunity to tackle many practical problems that are currently infeasible to solve, even with the most powerful classical supercomputers." (41)

> "Quantum technologies will make it possible to do things that simply cannot be done today." (7)

These different modes of change indicate that the scope and depth of quantum-induced change is framed in different ways across the discourse: projections range from incremental improvements to more speculative accounts of fundamentally transforming existing systems. I account for these differences when I synthesise the results of this section into four main visions of quantum computing.

3.3 Four visions of quantum computing

Building on the findings presented in Section 3.1, I identify four visions of quantum computing that emerge from across key actors in the field. I start from the overarching observation that quantum computing is consistently envisioned as bringing about change, usually framed positively in terms of progress. The four societal domains in which this change is envisioned serve as the basis for identifying different visions of quantum computing's societal significance: the social, economic, scientific, and geopolitical sphere. Within each domain, the change-making capacity of quantum computing is framed in a particular way – respectively, as a contributor to the public good, a scientific tool, an economic asset, and a geopolitical advantage. These framings constitute the four visions of the future of quantum computing.

Each vision is anchored in a 'value vocabulary': the implicitly or explicitly value-laden terms that actors use to frame quantum computing (e.g. as a scientific tool). As noted in the previous section, there is variation in the ambitions and scope of quantum-induced change. This is accounted for by distinguishing between narrower, more modest articulations of each vision and broader, more speculative forms.

I emphasise once more that these visions are *synthetic*: they result from aggregating patterns across actors and cannot be attributed to any single actor or actor group. They are also not mutually exclusive. They represent recurring ways in which the future of quantum computing is imagined across actors. I discuss each of them below.

3.3.1 Quantum computing for the public good

This mission-driven vision situates quantum computing within the social sphere, portraying it as a tool for addressing collective challenges – from healthcare and food security to sustainable energy and inequality. It frames quantum computing as a 'force for the public good', significantly contributing to societal goals and solving the grand challenges of society. The value vocabulary it appeals to includes terms such as "societal benefit", "wellbeing", "improvement", and "progress". In its more modest articulation, the technology is expected to contribute targeted solutions to specific societal goals. In its more expansive form, it is envisioned as a disruptive, general-purpose technology that is essential for navigating complex global problems and with

the potential to radically alter society at large. The quotes below are representative for this vision:

> "We believe that quantum computers will be uniquely capable of solving some of the world's most pressing issues." (26)

> "We're on a mission to build the world's most powerful computers to help solve humanity's most important and pressing problems." (31)

> "Our vision is to enable the UK to solve some of the most complex and challenging problems facing society by harnessing the potential of quantum computing." (40)

3.3.2 Quantum computing as scientific tool

Projected onto the scientific sphere, quantum computing is primarily envisioned as an epistemic instrument, extending or surpassing the capabilities of tools grounded in classical physics. It reflects a framing in which quantum computing serves the advancement of science – in particular physics, chemistry, and materials science – and the pursuit of fundamental knowledge. This vision is underpinned by value-laden terms such as "curiosity", "breakthrough", "discovery", and "progress". In the modest variant, quantum computing is portrayed as a new research apparatus; in more speculative variants, it is imagined as enabling paradigm-shifting discoveries. This vision is illustrated by the following quotes:

> "A critical driver for further advancement of the field and a top priority for Switzerland is curiosity driven research and innovation." (12)

> "Quantum computers will help us to get a better grasp of [quantum mechanics] by allowing us to simulate quantum phenomena that would otherwise just be too hard to study. […] These simulations could lead to advancements in chemistry, materials science, and fundamental physics." (60)

3.3.3 Quantum computing as economic asset

In a third vision, quantum computing is envisioned as driving change within the economic sphere. This vision frames quantum computing as a business opportunity, a key driver of economic growth and industrial competitiveness. It is portrayed as both a catalyst for innovation and a platform for new business models, with applications across a wide range of sectors. This vision appeals to term like "growth", "innovation", "value creation", "competitiveness", and "opportunity." In its more modest version, the focus is on incremental innovation and process optimisation, while in its more speculative form it anticipates a technological revolution underpinning a new economic paradigm. The following quotes are illustrative for this vision:

> "[The goal is] achieving a fully commercial quantum economy." (38)

> "Quantum computing is seen as a transformative paradigm for digital economies and societies." (37)

> "With the first paradigm-shifting advance since the integrated circuit, quantum computing is poised to transform every global market. [...] The first companies to use quantum computing to realize improvements in cost, speed, or accuracy over the best possible classical solutions could see significant changes to their bottom line, save years of development time bringing new products to market, or break ground in new industries yet to be imagined." (31)

3.3.4 Quantum computing as geopolitical advantage

In the geopolitical sphere, quantum computing is framed both as an opportunity to be seized and as a threat to be managed. More than the other visions, this one foregrounds the dual-use character of quantum computing, combining expectations of economic and military advantage with awareness of significant cyber security risks. Quantum computing is thereby positioned as a strategic asset tied to security, influence, and international competitiveness, conferring geopolitical advantage on those who possess it. Its value vocabulary spans terms such as "sovereignty," "leadership," and "opportunity," alongside "risk", "threat", and "security." In narrower articulations, the focus lies on cyber security threats; in its broader shape, the vision is embedded in narratives of a global race for technological dominance. Illustrative quotes include:

> "For Austria as a business location, it is important that companies invest early in future technologies and score points with innovative products 'made in Austria'. This applies in particular to quantum technologies, which hold enormous potential." (2)

> "Now we have another opportunity to be among the world's leading forces in a new technological revolution: quantum technology. [...] Quantum technology can also strengthen Finland's national security in several ways in both cyber security and defence." (8)

> "[Q]uantum technology is closely tied to immense momentum in the global innovation race. Germany and Europe are challenged to significantly intensify their efforts in the coming years. How we deal with the topic of quantum computing is indeed also a question of technological sovereignty in global competition with strong rivals, especially in Asia and the USA." (11)

> "Quantum technologies have potentially revolutionary and disruptive implications, which can degrade the Alliance's ability to deter and defend. Quantum technologies are therefore an element of strategic competition." (42)

Table 2 summarises the visions and their analyses.

Table 2. Four visions of quantum computing and its societal significance

Vision	*Sphere*	*Dominant framing*	*Scope*	*Value vocabulary*
Quantum computing for the public good	Social	Quantum computing is framed as a *contributor to societal goals*	*Narrow*: targeted, niche solutions *Wide*: general-purpose enabler	societal benefit; wellbeing; improvement; progress
Quantum computing as a scientific tool	Scientific	Quantum computing is framed as an *epistemic instrument*	*Narrow:* new research instrument *Wide*: paradigm-shifting discoveries	curiosity; breakthrough; understanding; discovery
Quantum computing as an economic asset	Economic	Quantum computing is framed as *innovation driver and business enabler*	*Narrow*: incremental innovation *Wide*: economic revolution	growth; innovation; value creation; competitiveness; opportunity
Quantum computing as a geopolitical advantage	Geopolitical	Quantum computing is framed as *strategic and security asset*	*Narrow:* cyber security threat *Wide:* global race for leadership	sovereignty; leadership; opportunity; security; dual-use; threat

4. Ethical analysis of visions of quantum computing

In this section, I analyse the ethical dimensions of the visions of quantum computing outlined in the previous section. I seek to uncover patterns and presuppositions embedded within and appearing across these visions, and reflect on their ethical aspects. Rather than assessing each vision in isolation, I analyse them collectively as expressions of the contemporary discourse on quantum computing. This involves interrogating how these visions frame desirable futures, delineate problems in need of resolution, and orient trajectories of technological development.

Before proceeding to the analysis, two points are worth noting. First, the analysis is exploratory in character. It seeks to flag ethically salient aspects of current visions of quantum computing, identifying where ethical concerns already arise or may emerge.

The resulting 'ethical heatmap' thus serves as an agenda outlining issues that warrant closer scrutiny, rather than as a comprehensive normative assessment.

Second, the aspects I highlight are not claimed to be unique to quantum computing. Rather, most of them are familiar issues resurfacing from previous NEST-ethics discourses. I contend that issues need not be specific to quantum to be relevant to quantum (see also Grunwald, 2005, for a similar argumentation in the context of nanotechnology; and Chapter 1, Section 2.2). Following Shelley-Egan and de Jong (2025), I take seriously the importance of continuities in ethical discussions about new and emerging technologies, as attending to them serves as a useful heuristic for ethics and allows us to learn from previous debates rather than treating each new technology as ethically unprecedented. By identifying how recurring issues manifest in the new technological context of quantum computing, its emerging ethical discourse is located within the broader tradition of NEST-ethics.

As outlined in Section 2, this analysis is structured across three analytic levels: the substantive level (content), the discursive level (framing), and the agential level (authors).

4.1 Content (substantive level): techno-optimism and technosolutionism

4.1.1 Techno-optimism

As the discussion of the value-oriented language mobilised in the visions shows, a pervasive *techno-optimism* characterises the contemporary discourse. One might be tempted to see such optimism as implicit in the notion of a vision, understood as a desirable and pursue-worthy future. However, this would be too simplistic. Visions of the future can also include negative imaginaries, articulating outcomes to be avoided and developments that should be prevented. Our analysis nonetheless shows that optimism strongly dominates the discourse. When left largely unquestioned, this can give rise to an overly confident, unreflexive or outright naïve understanding of quantum computing's future. I return to this concern of an imbalanced discourse in the next section.

The techno-optimism in visions of quantum computing takes two forms: optimism about its positive impact on the world (impact optimism), and optimism about the feasibility of realising this impact (feasibility optimism). I start with the latter.

Feasibility optimism cuts across all four visions. It reflects a dual confidence in both the technical and social feasibility of quantum computing. The innovation process is often portrayed as linear and inevitable, with the realisation of practical quantum computing framed as merely a matter of time:

> "It is no longer 'if' but 'when'." (1)

This optimistic, at times even deterministic frame rarely displays epistemic humility or engage seriously with the profound uncertainties[9] surrounding the development trajectory of quantum computing. Instead, these are treated as practical hurdles that will be overcome in time. Such neglect is facilitated by the abstract, often vague language used to articulate projections, which risks downplaying substantial scientific and technical challenges while fuelling hype and inflated expectations.

The social feasibility of quantum computing – that is, its capacity to function effectively within real-world contexts – likewise tends to be taken for granted. Although some visions acknowledge enabling conditions such as education, skills development, or infrastructure, these are typically framed as items to be addressed. The underlying assumption remains that, *once* the right conditions are in place, quantum computing *will* deliver on its promises.

This assumption warrants ethical scrutiny. Without denying the technology's potential, it is ethically important to recognise profound uncertainties involved, not least because they are relevant for responsible decisions about investment and broader policy (cf. Nordmann & Rip, 2009; see also Chapter 8). Yet across all four visions – albeit to different degrees depending on their speculative scope – such uncertainties are seldom mentioned, let alone meaningfully acknowledged. This raises questions about the justification of substantial financial and intellectual commitments to a technological future that remains fundamentally uncertain. I take up these questions in Chapter 8, when I scrutinise the concrete precautionary measures taken in the anticipation of quantum's threat to cyber security.

When it comes to impact optimism, the visions display a strongly positive view of the change associated with future quantum computers. Quantum computing is frequently cast as playing a key role in realising desirable futures. On this basis, the visions can be characterised as techno-optimistic (Danaher, 2022; Königs, 2022). While modest forms of optimism are not necessarily problematic, stronger forms that frame a technology as necessary or sufficient for achieving desirable futures invite ethical scrutiny.

A critique of such future-oriented impact optimism is articulated by Wilson (2017), who argues that techno-optimism becomes contestable in the absence of reliable probability estimates. He writes:

[9] At present, the feasibility of large-scale quantum computers is the 'million-dollar question' in the field. Expert forecasts indicate a broad consensus that large-scale quantum computers are likely to emerge within the next thirty years (Mosca & Piani, 2026). However, these projections have remained relatively stable in recent years, which suggests that, despite progress, the gap between current capabilities and practically relevant quantum computing remains substantial.

> "It is fine to be optimistic when probabilities are explicitly known to be in our favor. But when they are not known, and cannot possibly be, as in the case of our evaluation of the techno-scientific future, does reason not call on us to remain neutral?" (p. 349)

Given the substantial uncertainties surrounding the development of quantum computing, and the historically 'bumpy' nature of technological trajectories and the unpredictability of their societal impacts (Grunwald, 2026), the prevalence of techno-optimism in contemporary quantum computing discourse warrants critical examination. It becomes ethically concerning in particular when it informs or steers decision-making about the technology (again, see Chapter 8 for a concrete discussion of this issue).

Beyond this epistemic critique, techno-optimism can also be questioned on more substantive grounds: are its claims about positive societal impact of technological innovations actually plausible? This question brings us to the notion of technosolutionism.

4.1.2 Technosolutionism

Closely linked to this optimism is a pronounced tendency towards *technosolutionism*, understood here as the belief that technology offers effective solutions to complex societal problems (Sætra, 2023). Across the identified visions, quantum computing is repeatedly framed as a solution-generating technology, capable of – and often presented as key to – addressing not only technical challenges but also broader social and economic problems. This is consistent with other analyses in the literature, especially around the concept of "Quantum for Good" (Shelley-Egan, 2025; see also Chapter 7).

Technosolutionism has been widely critiqued for reducing complex, multidimensional social problems to narrowly defined technical issues (Morozov, 2013; Broussard, 2019; Siffels & Sharon, 2024). It risks sidelining structural and moral dimensions in favour of quick 'technofixes', not seldom generating new problems (Nydal et al., 2024; Grunwald, 2026). At the same time, framing a technology as contributing to the solution of a problem is not in itself ethically problematic. This would deny the defining problem-solving aspect of technologies (Chapter 2, Section 2.2). The ethical concern here is more nuanced and context-dependent (Ruser & Machin, 2016; Nydal et al., 2023; Sætra & Selinger, 2024).

Beyond concerns about technology-first approaches that privilege technical interventions over normative and political considerations, and about effectiveness, given that technosolutions may leave the roots of the problems they purport to address untouched (Grunwald, 2026), Nydal et al. (2023) highlight the risk of *de-moralisation*: technological solutions may obscure or displace the moral issues underlying the problems they aim to address. It is in particular this problem of moral opacity that can render technosolutionist framings ethically problematic.

Viewed through this lens, the vision of quantum computing as serving the public good becomes particularly salient, especially in its stronger forms that present the technology as essential for addressing society's 'grand challenges'. By explaining quantum computing's contribution primarily in terms of increased computational power, this vision tends to tacitly reframe 'society's hardest problems' as computational problems. Such framing has ethical consequences: it shapes what is recognised as a problem in the first place and narrows the range of responses considered appropriate – echoing that 'for those with a hammer, every problem appears as a nail' (see Chapter 7 for a more detailed discussion).

Whether such framings are ethically problematic depends on context and, ultimately, on how they shape concrete interventions and their consequences. Technosolutionism is therefore flagged as an ethically salient pattern in contemporary quantum computing discourse that warrants further scrutiny, in particular with regard to how it consolidates and translates into practice (Nydal et al., 2023).

4.2 Framing (discursive level): the good lies in the future and requires innovation

4.2.1 Promissory futurism

Shifting the analytic focus from the 'what' to the 'how', it stands out that the visions of quantum computing are consistently framed through a *future-oriented discourse*: the future is presented as the primary site in which societal prosperity will be realised. Within this framing, quantum computing is positioned as a key enabler of that desirable future.

At first glance, this comes not as a surprise: if visions are understood as projections of how the future might or should look, they are inherently future-oriented. Yet the strongly instrumental framing of quantum computing – as a means to clearly desirable ends – warrants closer scrutiny. When such a direct link between a technology and promised benefits comes to dominate the discourse, other ethically relevant dimensions risk being obscured, including potential downsides, the possibility of unintended consequences, value conflicts, and opportunity costs. If innovation is prefigured as inherently beneficial, then ethical reflection may be limited to questions of *how* to steer its development responsibly, rather than whether the innovation itself is desirable.

Moreover, locating 'the good' primarily in a distant future may shift ethical attention away from the present. It can defer responsibility and distract from more immediate interventions or alternative courses of action. While this concern overlaps with critiques of techno-optimism and technosolutionism, it highlights a distinct temporal dimension: technology is not only framed as an effective problem-solver, but its promised solutions are consistently projected into a speculative future.

The prevalence of such promissory futurism in the contemporary quantum computing discourse – where beneficial impacts are largely presupposed and framed as something that merely needs to be unlocked – warrants attention. Not only may it indicate an imbalanced discourse, but also raises ethical concerns about speculative overreach (Nordmann, 2007), which may undermine the quality of public and ethical debate (Chapter 4).

4.2.2 Moral legitimation

Closely related to promissory futurism is a discursive pattern of *morally legitimising innovation through anticipated social and economic benefits*. Projected contributions of quantum computing – to society, the economy, science, and national positioning – are mobilised to justify continued financial and intellectual investment. Such a logic is characteristic of innovation contexts and is not inherently problematic. However, like other recurring patterns and presuppositions, it becomes contestable beyond a certain point.

Ethical concerns arise when desirability frames are mobilised whose plausibility is uncertain or insufficiently substantiated – with 'sufficiently substantiated' being debatable. In other words, promising 'rose gardens' becomes problematic when such promises cannot be made plausible, yet still serve to attract or justify the allocation of substantial resources. In such cases, a form of spurious or *pseudo-legitimation* may occur, which is ethically problematic. Whether this dynamic is at play around quantum computing requires closer examination of individual claims.

This issue of legitimation is strongly tied to questions of discursive agency, asking *who* articulates these promises. The next section takes up this question.

4.3 Authors (agential level): interest shaped narratives

4.3.1 Stakeholder bias

The question of authorship adds a further dimension of ethical significance to the patterns and presuppositions discussed at the previous levels. *Who* articulates visions of quantum computing matters, because discursive power can translate into agenda-setting and action, directly or indirectly. It makes the difference between techno-optimism being merely naïve or potentially harmful. In other words, the ethical relevance of visions partly stems from the fact that they have authors.

As discussed in Section 3, I did not examine how societal actors such as civil society organisations or citizens imagine the future of quantum computing, for reasons explained there. Instead, I focused on actor groups that actively articulate visions of what the future of quantum computing will or should look like, and thereby currently shape the dominant discourse. These actors are not only discursively visible but also

institutionally powerful, and their visions play an active role in steering the development of quantum computing. This power renders their visions particularly ethically salient. I here reflect on two features of the current discourse and their potential ethical implications.

First, the identified visions are largely authored by actors with *direct institutional or strategic stakes* in the advancement of quantum computing. For example, while the 'quantum for societal good' vision rhetorically centres on benefits for "humanity", its principal authors – national governments – also emphasise the geopolitical and strategic value of maintaining leadership in quantum technologies. Similarly, the economy- and science-oriented visions are shaped by institutions whose legitimacy, funding, or influence may depend on projecting quantum computing as a major opportunity. These overlapping stakes reinforce a strong desirability frame and contribute to a discourse in which more critical questions – such as those concerning energy use, reliance on scarce materials, or the need for non-technological forms of change – tend to receive limited attention.

That visions are shaped by interests is not in itself surprising. Ideas about how the future should (or should not) look are closely tied to what different actors consider desirable, often from their own perspective. However, if the discourse remains largely shaped by a narrow set of elite actors, a correspondingly narrow – or at least unchallenged – set of interests risks dominating it, and the ethical concern arises that broader societal stakes are sidelined in setting innovation agendas. From an ethical perspective, a more plural and inclusive discourse would therefore be preferable.

4.3.2 Moral ambivalence

A related but more implicit issue concerns the value tensions embedded within these visions, most notably between sustainability and economic growth (Atladóttir et al., 2025). Economic growth is often framed both as a goal and as a justification for quantum innovation, yet this may stand in tension with sustainability considerations, particularly in light of the material and energy demands associated with the technology. The different values mobilised in the visions and the tensions between them raise questions not only about the compatibility of different visions, but also about internal inconsistencies within individual visions.

Such ambiguity or moral ambivalence is common in discussions of new and emerging technologies (Shelley-Egan & de Jong, 2025). While Shelley-Egan and I explain this ambivalence in terms of framing a technology as novel versus business as usual, the findings here suggest that it also manifests in the simultaneous appeal to potentially competing values in visions of quantum computing. This is not ethically problematic in itself. However, when such value tensions remain implicit or unaddressed, they obscure underlying trade-offs and normative choices. Making these tensions visible

therefore matters ethically, as it opens space for explicit reflection and contestation over what kind of progress is being pursued, and for whom.

Table 3 summarises the ethical analysis.

Table 3. Overview of the findings of the ethical analysis of quantum computing visions

Level of analysis	*Feature*	*Potential concern*
Content (substantive level)	**Techno-optimism**	Downplaying uncertainties around the development of practically relevant quantum computers may compromise responsible decision-making
	Technosolutionism	Presenting quantum computing as a solution to social problems may obscure the social complexity of those problems and their sociopolitical and ethical aspects
Framing (discursive level)	**Promissory futurism**	Framing quantum computing as enabling 'the good in the future' risks speculative overreach, which may negatively affect the quality of public and ethical debate
	Moral legitimation	Legitimising innovation agendas by appealing to future benefits of quantum computing without substantiated plausibility risks pseudo-legitimation
Authors (agential level)	**Stakeholder bias**	Visions of quantum computing shaped by a narrow set of powerful actors may sideline broader societal interests
	Moral ambivalence	Unacknowledged tensions between values in different visions of quantum computing obscure ethical trade-offs and keep the ethical debate about what counts as progress implicit

To a high extent, this analysis reveals similarities with previous debates about system technologies such as nanotechnology and artificial intelligence (Swierstra & Rip, 2007; WRR, 2021). This provides openings for transferring insights from these previous debates to the contemporary discourse on quantum computing.

5. Conclusion

Visions of technology do not merely reflect anticipated futures, but are value-laden narratives that actively shape technological trajectories. For this reason, projections of the future of quantum computing warrant ethical scrutiny. Moreover, studying visions of future technologies allows for early-stage ethical engagement without relying on speculative or potentially contested assumptions about how the technology will develop.

In this study, four main visions of quantum computing were identified, sourced from across key actors in the field. Each of these visions casts the societal relevance of quantum computing in a distinct role: as a contributor to the public good, a scientific tool, an economic asset, or a geopolitical advantage. Taken together, these visions form a powerful discursive force that shapes how quantum computing is framed and developed.

Analysing these visions through an ethical lens brings several observations and concerns to the fore. First, a pervasive techno-optimism embedded in these projections was noted, which tends to obscure the uncertainties and challenges inherent to quantum computing's development. This optimism is closely tied to a technosolutionist logic – one that positions quantum computing as a key enabler of diverse societal goals, potentially backgrounding the sociopolitical dimensions of these goals and the ethical implications of pursuing them through technology.

Second, I highlighted the promissory language through which these visions justify current innovation efforts by appealing to speculative future benefits. While this is a familiar feature of innovation discourse, it becomes ethically significant in the context of quantum computing's early stage and the considerable technical and conceptual hurdles that remain before such promises can be realised, potentially collapsing into pseudo-legitimation.

Third, I observed that the authors of contemporary visions of quantum computing all have direct stakes in its development. While this is not surprising in itself, the fact that the discourse is currently dominated by a small, powerful elite raises concerns that broader societal interests may be neglected. Moreover, different authors appeal to different values in their visions, which sometimes come into tension with one another. This moral ambivalence highlights potential conflicts between visions and raises questions about their compatibility and practical feasibility. Such observations also point toward a need for more detailed scrutiny of specific risk claims associated with quantum computing. These too are value-laden and shaped by the perspectives of those who advance them. Chapter 8 takes up that thread, assessing the cyber security threat posed by quantum computing as a specific kind of vision.

These aspects are not necessarily unique to quantum computing; they reflect recurring patterns in ethical discussions about new and emerging technologies. By identifying whether and how these issues resurface in the context of quantum computing, our analysis sets an agenda for more focused ethical enquiry into each of them. It also underscores the value of attending to continuities with previous debates – such as those on nanotechnology – which allows us to draw on, and learn from, earlier ethical discussions.

With this ethical vision assessment, I aim to make a threefold contribution. First, to the emerging field of quantum ethics and the broader discussion of responsible

innovation in the quantum realm. This analysis functions as an ethical heatmap of potential concerns arising from how the future of quantum computing is currently envisioned. Rather than providing an exhaustive ethical evaluation of each vision, this chapter highlights patterns and presuppositions that warrant closer scrutiny, thereby identifying entry points for further ethical enquiry and public reflection. These concerns should be seen in light of the largely absent broader societal perspectives in current discourse. A key avenue for future research is therefore to explore public perspectives on quantum computing in relation to the 'vanguard visions' examined here.

Second, this study contributes to methodological discussions on applying a hermeneutic or meaning-oriented approach to ethics, especially for technologies still in an early, conceptual stage. I propose a structured method for identifying technological visions from textual materials, and for interrogating their ethical dimensions. By explicitly incorporating ethical analysis as a distinct stage of vision assessment, technological visions are foregrounded as ethically salient phenomena and illustrate how a meaning-oriented approach to ethics of early-stage technology can be pursued (see Chapter 5).

Third, this analysis situates the visionary communication currently surrounding quantum technologies within a broader lineage of similar waves associated with other new and emerging technologies; earlier ones like nanotechnology, and current ones such as synthetic biology, gene editing, and artificial intelligence. These technologies, too, have been, or still are, accompanied by visionary communication, at times reaching the scale of powerful hype. Common threads include technosolutionism, the legitimisation of present-day investments and efforts through promises of a better future, and the particular stakes and interests of researchers and research managers. In-depth comparative studies of techno-visionary communication and its ethical dimensions across these different waves of new and emerging technologies would therefore be a promising avenue for better understanding their commonalities and differences.

Finally, having argued that the uncertainties surrounding quantum computing development are relevant for ethical analysis, this insight also applies to the current study. How can engaging with visions of a technology whose feasibility remains deeply uncertain be justified? As Nordmann and Rip (2009) point out, the relevance of ethical analysis partly depends on the plausibility of the projections it addresses. While this holds directly for outcomes-oriented approaches focused on technological impacts, the ethical relevance of engaging with visions depends less directly on their plausibility. Since visions mobilise resources, shape expectations, and influence what eventually becomes possible, they merit ethical scrutiny regardless of the current state of technical feasibility. In this sense, our method of ethical vision assessment offers a distinctively anticipatory approach to ethical engagement with emerging technologies.

The next chapter examines more closely one of the dynamics flagged here: technosolutionism in the quantum discourse.

Chapter 7

Technosolutionism and the Case of Quantum for Good

An earlier version of this chapter was published as: de Jong, E., & Shelley-Egan, C. (2026). Quantum for Good through a lens of technosolutionism: A critical-constructive view. Journal of Responsible Technology, 25, 100157. DOI: https://doi.org/10.1016/j.jrt.2026.100157

Abstract

'Quantum for Good' (Q4G) has emerged as a guiding rationale for quantum technology development, emphasising its potential to address societal challenges and serve the public good. Through this normative ambition, Q4G echoes the language of responsible innovation. At the same time, using technology to address complex social problems has been criticised as 'technosolutionism': the belief that technology is an effective means to solve non-technological problems. In this chapter, I examine whether and how this critique applies to Q4G. The chapter adopts a principally neutral perspective on technosolutionism, recognising both the potential and limits of technological solutions, and use this as a critical-constructive lens to assess the normative underpinnings of Q4G. Specifically, I develop a set of guiding questions to scrutinise the ethical dimensions of proposed technosolutions and apply this framework to Q4G. I argue that, to live up to its commitment to societal good and to avoid becoming a legitimising narrative for existing innovation agendas or reducing complex issues to technological challenges, Q4G must remain attentive to the ethical and sociopolitical dimensions of the problems it targets and the solutions it advances.

1. Introduction

As noted in Chapter 1, although most quantum technologies remain at an early stage of development and their practical applications are still largely speculative, a discourse about their potential societal value is already taking shape. The previous chapter mapped the visions shaping this discourse. This chapter focuses on one issue flagged there as warranting closer scrutiny: the tendency to frame technology as an effective solution to complex societal problems, also known as 'technosolutionism'. This logic is often accompanied by another dynamic flagged earlier: a strong optimism about technological development itself, that is, about whether a technology will materialise and perform as envisioned. In this chapter, I examine both technosolutionism and the techno-optimism that typically underpins it in the emerging discourse around quantum technologies, in particular under the banner of 'Quantum for Good'.

Alongside narratives that position quantum technologies as a scientific, economic, and geopolitical asset (Chapter 6), *Quantum for Good* (Q4G) has emerged as a prominent framing and rationale for quantum technology development. It refers to the leveraging of quantum technologies to address pressing societal challenges and to contribute to the public good, however defined. These aspirations are often articulated through alignment with global agendas such as the UN Sustainable Development Goals (SDGs) and, importantly, through a focus on identifying concrete quantum use cases that demonstrate societal benefit.

As such, Q4G embodies a normative ambition, founded on assumptions about the desirability of technological solutions and the instrumentalisation of quantum technologies in service of societal values. In this respect, Q4G echoes the language of responsible innovation (e.g. Von Schomberg, 2013) and appears to answer calls for such approaches in the quantum domain (e.g. Gasser et al., 2024) – although the extent to which it genuinely embodies these commitments warrants closer examination.

Despite its increasing prominence, Q4G has received little critical scrutiny. While the concrete manifestations of Q4G remain nascent, it is important to examine the assumptions underpinning such efforts from the outset. Framed as delivering societal benefits, these initiatives call for ethical assessment, particularly when public resources are involved. Moreover, if the Q4G frame gains widespread endorsement, it could come to shape the impact pathways of quantum technologies in society. Thus, an explicit ethical appraisal of the notion is required to better understand its underlying assumptions and possible ramifications.

At the heart of Q4G and its orientation towards finding use cases lies a belief in the transformative potential of quantum technologies. This focus on the problem-solving capacities of technology links Q4G to technosolutionism: the notion that we can and should use technology to solve the challenges we face (Sætra, 2023). Technosolutionism has often been criticised for its reductive logic and the implicit innovation agendas it

serves. The concept is therefore frequently treated pejoratively. Drawing on recent scholarship, I adopt a more nuanced and diagnostic use of the term, recognising both the potential and limits of technological solutions. Rather than dismissing technological initiatives outright, I employ technosolutionism as a – critical yet constructive – analytical lens to evaluate Q4G as a normative project.

Using the concept of technosolutionism in unpacking Q4G allows us to place front and centre the sociotechnical conditions in which quantum technology development is taking place. It enables a reflective interrogation of how technologies are framed as solutions, and how societal problems may be reshaped – or narrowed – to fit those solutions (Nydal et al., 2023). Examining Q4G through this lens draws attention to the risks posed by technosolutionist tendencies while also identifying opportunities for a more reflexive and responsible approach to Q4G. Thus, rather than evaluating the potential impacts of quantum technologies – the typical focus of ethical analysis – I focus on the reasoning underpinning Q4G as a motivating rationale and shaping force in their development.

I begin by introducing the emergence of Quantum for Good as a normative ambition (Section 2), examining its core assumptions and its focus on identifying quantum 'use cases'. I then synthesise existing accounts of technosolutionism (Section 3) and distil from them a nuanced view that can be mobilised for assessing technosolutions. Building on this perspective, I develop four guiding questions for the ethical assessment of specific technosolutions (Section 4). This framework[1] is then applied to evaluate the project of Q4G, with particular focus on efforts to align quantum technologies with the UN Sustainable Development Goals (Section 5). I conclude with an overall assessment of Q4G and suggestions for strengthening it as an ethically robust normative ambition for societally attuned innovation (Section 6).

2. The emergence of Quantum for Good

The year 2025 has been designated by the United Nations as the 'International Year of Quantum Science and Technology'. This initiative celebrates "100 years of quantum progress" and aims to "raise public awareness of the importance and impact of quantum science and applications on all aspects of life." (Quantum 2025, n.d.) The prevailing tone is one of optimism about a 'quantum-enabled future', which is not only associated with significant economic benefits and scientific breakthroughs but also with societal progress. The perceived potential of quantum technologies to contribute to the

[1] I do not present this framework as a new addition to the landscape of RRI initiatives. I fully acknowledge that many such frameworks already exist. The aim is instead to analyse the emerging dynamics and rhetoric surrounding quantum innovation – dynamics I characterise (without dismissing them) as technosolutionist – and to structure the resulting ethical considerations into a set of guiding questions that support a critical-constructive evaluation of Q4G.

good of society and the initiatives actively pursuing this, is what has been popularised as Quantum for Good (Q4G). In this section, the emergence of the Q4G is examined as a normative ambition to harness quantum technologies for societal benefit (2.1) and I reflect on the role of use case development in shaping that ambition (2.2).

2.1 Quantum for Good as a normative ambition

Building on a discourse that has primarily emphasised the scientific, economic, and geopolitical promise of still-emerging quantum technologies, the phrase 'quantum for good' has gained traction in recent years (Vermaas & Mans, 2024; Shelley-Egan, 2025; Shelley-Egan & de Jong, 2025). The implicit reasoning seems to be that if quantum technologies could drive scientific progress, economic growth, and strategic power, they might also be mobilised for social benefit. And if we are to invest in quantum's anticipated scientific, economic, and geopolitical gains, we should also invest in quantum initiatives oriented towards societal goals.

The phrase 'quantum for good' was preceded by an association of quantum technology with ambitions including "benefitting humanity" (World Economic Forum, 2022), contributing to the "public good" (Roberson et al., 2021), and "advancing society" (Kop et al., 2024a, 2024b; Gasser et al., 2024). Over the past few years, the concept has featured in (semi)academic conferences, further solidifying its presence in discussions on quantum technology's societal role.

In 2024, the United Nations International Computing Centre (UNICC) organised an event explicitly centred on Q4G and its potential applications for sustainable development (UNICC, n.d. (a)). This was followed by a conference in The Hague, organised by the UNICC, the International Telecommunication Union (ITU) and Quantum Delta NL (QDNL), which explored the "transformative potential of quantum technology for addressing global challenges" (UNICC, n.d. (b)). That same year, 2024, the UNICC and QDNL launched a partnership under the banner of Q4G, initiating a global collaborative movement "dedicated to creating positive societal, environmental, and economic impact through quantum technology" (QDNL, 2024; Quantumforgood.eu, n.d.). If 2025 is the official Year of Quantum, then 2024 marked the launch of Q4G.

Broadly speaking, Q4G frames quantum technologies as a tool with which to address pressing societal challenges and the public good. The term suggests a commitment to ensuring that quantum technologies are not merely developed for commercial or military applications but are also directed towards solving social and environmental problems. Q4G reflects a normative ambition to harness the emerging capabilities of quantum technologies – such as quantum sensing, quantum computing, and quantum communication – for beneficial purposes, often aligned with broader global goals like sustainability, healthcare advancements, and economic equity. I stress 'emerging' or

'potential' capabilities here, as, while there is solid scientific ground for anticipating such applications, their practical realisation currently remains constrained by substantial scientific and engineering hurdles.

Concretely, Q4G is often linked to the UN Sustainable Development Goals (SDGs), with the argument that quantum technologies could contribute to fields such as climate modelling (Goal 13: Climate Action), drug discovery (Goal 3: Good Health and Wellbeing), and secure communication in ways that promote societal wellbeing. The Open Quantum Institute – a multilateral governance initiative launched in 2024 – further institutionalises this agenda, explicitly promoting the development of quantum applications "for the benefit of humanity" (Open Quantum Institute, n.d.).

The Q4G narrative offers a counterpoint to the risk-based framing that often dominates discussions on emerging technologies. Q4G foregrounds the potential positive societal impact quantum technologies, reflecting a shift from – or extension of – purely scientific and economic motivations to a more values-driven approach in technological development. This can be seen against the backdrop of an emphasis on responsible research and innovation over the last two decades, focusing not only on mitigating risks but also on directing technology development towards beneficial societal outcomes. Precedents such as ethical AI and the ITU's *AI for Good* initiative demonstrate how emerging technologies can be framed in normative terms. Q4G seems to be following a similar trajectory.

Positioning quantum technologies as a potential solution to societal challenges – or at least as a significant contributing technology to solving societal problems – helps to justify investments in their development and creates a basis for broader societal support. While this carries the risk of Q4G being used to legitimise commercial quantum agendas (see Section 5), it also draws much of its current credibility from existing investment in quantum's economic and strategic potential, creating a mutually reinforcing dynamic.

Setting the ambition of 'quantum for good' can guide current efforts in quantum research and development and nudge them in the direction of certain application domain. However, it remains an open question as to what exactly the 'good' in Q4G entails and whose 'good' it serves. In practice, this question is often addressed – implicitly – through the active search of 'use cases,' to which I turn next.

2.2 In search of 'use cases'

From the outset, Q4G has been presented as an effort to identify societally beneficial 'use cases'. The term use case originated in software engineering, where it referred to the technical requirements and user interactions involved in operating a system (Jacobson, 2004). Initially, use cases functioned as structured descriptions of user journeys, instrumental in designing interfaces and drafting user manuals. Over time,

the concept expanded into business modelling, where it came to include strategic considerations alongside technical specifications.

In more recent years – particularly in the context of emerging technologies such as artificial intelligence – the term has been used to denote targeted applications designed to address specific business challenges (e.g. Chui et al., 2018). Within the domain of quantum technologies, and especially under the banner of Q4G, the notion of use case has widened further. It no longer refers solely to existing uses or near-term applications. Instead, it increasingly appears to function as a way of identifying prospective opportunities and articulating normative goals that drive mission-oriented innovation, often reaching beyond the business realm into broader societal aspirations.

The active search for use cases has become a central theme in Q4G efforts. For example, the Open Quantum Institute explicitly aims to "realise the full potential of quantum computing by accelerating the development of use cases relevant to achieving the SDGs, leveraging the combined forces of researchers, developers, entrepreneurs, and United Nations and large NGO innovation cells" (Open Quantum Institute, n.d.). Similarly, the two Q4G conferences mentioned in Section 2.1 prioritised identifying quantum use cases.

The search for Q4G use cases can be understood as a specific operationalisation of its broader normative ambition oriented towards developing technological responses to societal challenges. This approach reflects a logic often described – and criticised – as *technosolutionism*. The following section discusses this logic in greater depth.

3. Technosolutionism: from pejorative term to critical lens

To examine whether and how critiques of technosolutionism apply to Q4G, a comprehensive account is needed of what these critiques entail. In what follows, I synthesise the state-of-the-art discussion about technosolutionism. I begin by outlining the main critical perspectives (Section 3.1), before turning to more nuanced approaches that treat technosolutionism not as inherently good or bad, but as a context-dependent phenomenon warranting critical evaluation (Section 3.2). Beyond providing an overview of the debate, this discussion highlights the conditions under which a technosolution may become ethically problematic. Section 4 builds on these insights to propose a set of questions to guide such evaluations.

3.1 Critical accounts of technosolutionism

In the face of complex societal challenges, technological innovation is often promoted as the path forward. Indeed, we 'innovated' ourselves out of the COVID-19 pandemic through the development of vaccines. Yet debates about the use of AI and Big Data in the development of apps for tracking infections and in pandemic control highlighted

the controversies surrounding technological responses to social problems (Siffels & Sharon, 2024).

Technological solutions to social problems have drawn renewed scrutiny against the backdrop of new technological advancements such as Artificial Intelligence and geoengineering (O'Neil, 2017; Broussard, 2019; Coeckelbergh & Sætra, 2023; Sætra, 2023; Sheikh, Prins & Schrijvers, 2023; Stein, 2024). The concept of 'technosolutionism' provides a useful lens for understanding these discussions. Technosolutionism centres on the idea that we should mobilise technology to solve the challenges we face (Sætra, 2023).[2] Thus technosolutionism always implies a degree of confidence in technology's potential to address complex societal issues, which, at times, may verge on overconfidence or even hubris (Johnston, 2018).

Criticism of technosolutionism is longstanding and comes from multiple angles. It ranges from fundamental objections to framing social problems as technical ones (Naess, 1973; Morozov, 2013) to more targeted concerns about the unintended harms and ethical blind spots of particular technologies (Broussard, 2019; Nydal et al., 2023; Siffels & Sharon, 2024). These strands of criticism reflect Scott's (2011) distinction between philosophical and practical critiques. Philosophically, technosolutionism is criticised for promoting a narrow mindset in which technological progress becomes the default mode of problem-solving, marginalising non-technological or social alternatives and excluding existing social practices as valid solutions.[3] Practically, it is challenged for failing to solve problems, creating new ones, or reinforcing flawed systems that ought to be transformed rather than superficially repaired.[4] Notably, these critiques

[2] Technosolutionism is often discussed alongside the notion of the 'technofix'. While some use the terms interchangeably, others have explicitly distinguished between them (Scott, 2011; Sætra & Selinger, 2024). In this chapter, I do not draw a sharp distinction between the two; I primarily adopt the term technosolutionism because it refers less to a specific technological intervention and more to an underlying logic or mode of thinking. This emphasis on process and orientation aligns more closely with the present analytical focus.

[3] Relatedly, Nachtwey and Seidl (2024) describe a "solutionist ethic" that has become central to the embedding of new technologies in society. This ethic is founded on the notion that solutionism comprises a new order of worth, in which value derives from solving social problems through technology.

[4] Scott illustrates these critiques through Timothy LeCain's historical study of pollution from copper smelting in the United States (LeCain, 2004). The problem in question was the removal of sulphur dioxide gas emissions – resulting from smelting processes - from the immediate environment, in addition to a conflict between mining companies and local communities. In order to address the problem, chemists adopted technological solutions, including, in one case, converting the pollutant into sulphuric acid to produce fertiliser – a seemingly elegant "win-win" fix. Yet this solution merely relocated the problem: the transformed and relocated environmental toxins – taken up in the fertilisers – had unforeseen environmental consequences. Furthermore, the 'fix' preserved the polluting industry itself, displacing more transformative alternatives. This example highlights how technosolutionism can manifest as narrow problem framings, unintended consequences, and the entrenchment of existing systems under the guise of innovation.

may overlap but are not always aligned: one can criticise a specific intervention without rejecting the problem-solving capacities of technology wholesale.

Philosophical and practical critiques of technosolutionism converge in concerns about how problems are constructed. A core critique of technosolutionism rests on the poor construction of the technosolutionist problem such that the problem definition follows the solution and is twisted to fit the solution, rather than the opposite way around (Siffels & Sharon, 2024). This directionality is especially problematic with respect to public policy making, where good problem construction implies careful deliberation and discussion of alternative solutions.

Moreover, this 'technology-first' approach is backward-looking, reflecting a "reformulation or recasting of an existing phenomenon as a problem (…)" (Siffels & Sharon, 2024, p. 4). For example, the existence of smart technology – such as sensors in the home that can collect data relevant to sleep cycles – can lead to sleep itself being recast as a 'problem' to be solved. By contrast, a forward-looking approach would start from the problem and then consider possible solutions, avoiding a narrow or forced definition of the problem based solely on the capabilities of existing technologies. This problem-first directionality is more likely to yield well-constructed problems, because the problem 'presents itself', emerging naturally from real-world relevance.

3.2 Nuanced views of technosolutionism

Several scholars have argued for a more open and context-sensitive view of technosolutionism (Ruser & Machin, 2016; Nydal et al., 2023; Sætra & Selinger, 2024) and to value concepts such as 'technosolutionism' and 'technofix' as tools for understanding the roles technologies play in modern society (cf. Scott, 2011; Ruser & Machin, 2016). Here, I bring them together to build a nuanced view of technosolutionism.

A first argument highlights the need to 'politicise' technosolutions: being attentive to the *social and political context* in which they emerge and are discussed. Ruser and Machin (2016), for example, call for scrutinising both the technofix narrative *and* its critique. They argue that this is necessary since both risk obscuring the deeper sociopolitical dynamics at stake and 'depoliticising'[5] democratic negotiation.

They illustrate this through two narratives about geoengineering. The post-growth narrative advocates for systemic change: reducing emissions and transforming high-consumption, fossil-fuel-based economies. The technofix narrative, in contrast, offers a 'second best' path: using technological innovation to address climate change without altering lifestyles. While these narratives differ in their vision of change, Ruser and

[5] I define depoliticisation as the process of framing issues, decisions, or policies as non-political, often by presenting it as purely technical or inevitable, and thus outside the scope of public debate.

Machin argue that both fall short in addressing the political questions embedded in technological development (p. 438). Notwithstanding questions about the feasibility of geoengineering itself, issues of authority and accountability – such as who has the power to approve and implement geoengineering experiments, and who bears responsibility for their outcomes – are often side-stepped. These concerns are especially salient given that only a few powerful, wealthy states are likely to possess the resources to deploy such technologies.

However, the 'techno-wariness' enveloped in the critique of the technofix narrative can also obscure questions of power: Who decides whether these technologies can be used or not? Moreover, rejecting technosolutions outright may risk denying resources or interventions to those who could benefit most (Wieners et al., 2023). In both acceptance and rejection of technological solutions, political judgment and social negotiation remain essential. In what follows, I will refer to these aspects as the 'sociopolitical dimensions' of technosolutions.

Ruser and Machin thus emphasise that technologies are always socially embedded and cannot be meaningfully accepted or dismissed outside political processes. As they write, "it cannot be instituted or rejected prior to political decision-making, but rather necessarily involves ongoing social and political analysis" (2016, p. 437). Thus, a truly critical (yet nuanced) approach to technosolutionism must rise beyond deep scepticism and instead engage with the sociopolitical dimensions of both innovation and its critique.

A second argument points to the need to critically evaluate the ambiguity of technofixes and their scope. This relates to the practical criticism that such fixes often preserve systems that arguably should be dismantled or transformed for better alternatives. Their ambiguity lies in the fact that their perceived success depends heavily on how, and by whom, success criteria are defined (see also Scott 2011, p. 225).

Scott warns of the need to be cognisant of and to manage expectations regarding what technological fixes can achieve, and the new kinds of problems they might bring about. At the same time, he argues that we should not dismiss their potential too readily. A technofix, he notes, "(…) may simply be the most appropriate response at the moment, or the best that can be done at this time" (p. 225), offering policymakers "(…) more options and additional means for addressing problems" (p. 225). In a similar vein, Sætra and Selinger (2024) argue that "addressing symptoms does not necessarily mean one is ignorant of or indifferent to the underlying causes" (p. 4.). A technofix may offer a palliative or 'band-aid approach' – and band aids are, at times, necessary or ethically appropriate in the short term. The point is, rather, to remain attentive to the full extent of the problem and the limits, in terms of time or scope, of the technosolution.

A third argument centres on the need to question the legitimacy of technosolutions. One version of this argument is closely related to the ambiguity argument and calls for

concepts like technofix and technosolution to be stripped of their normative baggage, and instead evaluated in terms of plausibility, appropriateness, and potential effectiveness. Treating these as normatively neutral concepts creates room for a more careful and less dismissive evaluation of technosolutions, resisting both cynicism and naïvety (Sætra & Selinger, 2024). Such a constructive approach allows for the possibility of a technological framing of a problem being a valid one.

While also concerned with the legitimacy of technosolutions, Nydal et al. (2023) highlight the normativity of technosolutionist approaches. Nydal et al. (2023) emphasise that technofixes can lead to a kind of 'moral desensitisation' that may come about as a result of technofixes hiding or backgrounding the specific conditions of their success:

> "Technologists' ability to background the conditions of their success calls attention to a key aspect of the techno-fix objection. The common techno-fix objection expresses an important concern of something being forgotten or not receiving appropriate attention. The real moral issues (…) are prevented from presenting themselves and are silenced in disturbing ways." (p. 3)

The authors thus emphasise that technofixes carry moral implications, and that these can be obscured when the social conditions enabling a technological solution are pushed into the background (p. 3). Their concern is not only that such conditions are forgotten, but that the underlying moral issues they embody are prevented from coming into view at all – hence, de-moralising the problem.

The issue here is one of technological opacity (Vallor, 2016) and its ethical consequences: the technofix can conceal the moral stakes of the issue to which it responds (e.g. climate change), thereby enabling continued, unreflective practices without fostering awareness or engagement with the underlying problem.[6] Nydal et al. (2023) conclude that such morally problematic opacity can be mitigated by placing the technology "in plain view": explicitly acknowledging the problem a technological intervention seeks to address, even if it does not structurally resolve it, and making visible the conditions that render the intervention possible (p. 14).

Ultimately, they argue that the moral legitimacy of tech-fixes must be evaluated case by case, to assess whether the technology silences morally relevant implications and facilitates problematic, unreflective behaviour, and to look for ways to sensitise people to the problem that is being backgrounded by the technological intervention.

[6] Nydal et al. (2023) demonstrate this with a description of a Spanish viticulture innovation in which technological efforts are made to solve the problem of alcohol levels in wine. This problem is brought about by the effects of climate change. Bioengineered yeasts mitigate the alcohol content of wine, while at the same time they "(…) may become another hidden techno-fix that operates behind the scenes and hides both the role of the technology and the seriousness of the climate change problem (…)" (p. 15).

Together, these arguments offer a more nuanced understanding of technosolutionism: recognising their potential to contribute to problem-solving, while also remaining alert to their limitations, risks, and sociopolitical dimensions. By shifting focus from blanket endorsements or rejections to questions of legitimacy, scope, and power, this discussion helps move beyond a binary view of technological fixes. This perspective opens up space for more reflective engagement with technosolutions: beyond asking what they solve, addressing questions about how, for whom, and at what cost. This, in turn, provides a foundation for developing practical guidelines for an ethical analysis of initiatives such as Q4G.

Following Scott (2011) and Ruser and Machin (2016), I contend that dismissing technological solutions – or positioning them as inferior to non-technological ones – misses two important points: the central role of technology in human life, and the need to assess whether a given technological intervention is effective and appropriate in context. The issue is not technology per se. Rather, problems arise when technology is treated as the solution to challenges that are fundamentally structural or social in nature. Structural issues typically require structural change, and social problems often call for social responses. Technology may well contribute to such responses, but it can also be misapplied as a superficial 'quick fix' that leaves underlying causes and the social dimensions of the problem largely untouched.

Building on this discussion, technosolutionism is here taken as a morally neutral concept: neither intrinsically beneficial nor inherently suspect. Whether a technosolution becomes ethically problematic depends on several factors, including its sociopolitical implications, the ambiguity of its success criteria, and the extent to which it obscures morally relevant aspects of the problem at hand. Mobilising these insights, the next section develops a set of guiding questions for examining these ethical dimensions in concrete cases, e.g. under the banner of Q4G.

4. Ethical assessment of technosolutions

Treating technosolutionism as neutral, as something that is not inherently problematic, does not mean suspending normative assessment. Instead, it means that its ethical status should be assessed case by case. A neutral starting point opens space for investigating whether a specific instance is ethically problematic, and if so, in what way. Drawing on the discussion in Section 3, this section proposes a framework of guiding questions to support exactly that investigation.

The framework is composed of guiding questions that can inform critical-constructive, ethical engagement with emerging or proposed technosolutions. These questions can guide *ex ante* analysis, supporting reflection *before* the technosolution has been conceived, as well as 'real-time' analysis, guiding ongoing evaluation once a concrete proposal has taken shape. While *ex ante* questions support anticipatory reflection and

norm sensitivity in the early stages of innovation, real-time questions allow for ongoing critical engagement as proposals take shape. All questions may also be applied *ex post*, although their anticipatory function diminishes at that stage.

4.1 Guiding questions for ethical scrutiny

From the critical accounts of technosolutionism – both philosophical and practical – I take problem construction as a foundational issue for ethical scrutiny. It is foundational because how a problem is framed shapes the range of solutions considered and the criteria by which success is measured. If the problem is poorly or misleadingly constructed, everything that follows – in particular, the proposed solution – rests on shaky ground. Following the discussion in the previous section, a particularly revealing dimension of problem construction is its temporal orientation: whether a technosolution frames the problem in terms of its root causes or its anticipated consequences shapes the range of solutions considered. This leads to the first question:

- *Problem construction:* How is the problem that a technosolution claims to address constructed? And in particular, does it have a backward-looking orientation (focused on addressing root causes) or a forward-looking one (focused on managing future consequences)?

From the nuanced accounts, I take the *sociopolitical dimensions* of technosolutions (and their critique), the scope of the problem (and its proposed solution), and the *legitimacy* of both the problem and its technological intervention as key additional concerns for scrutiny. This gives rise to questions about the motivations underlying a technological approach, the definitions of success, unintended consequences, and the possible obfuscation of broader moral or structural issues:

- *Sociopolitical conditions:* What motivations underlie the proposed approach and whose interests do they serve? Which criteria inform the proposed technosolution and according to which criteria of success? Who defines these criteria?
- *Scope:* What is the temporal scope of the technosolutionist intervention? Is it planned to be a temporary or longer-term solution? Have possible societal and environmental impacts been considered or identified?
- *Legitimacy:* Can the proposed technological approach be considered a plausible, appropriate, and potentially effective response, given the current context?[7] Which existing practices does this solution reinforce or leave unchallenged? Are efforts also made (elsewhere) to address more structural aspects of the problem?

[7] Assessing this involves both technical and non-technical considerations, which may require different forms of expertise.

Together, these questions constitute a framework for critically yet constructively examining technosolutionist approaches. Open to their potential value while remaining alert to their risks. The questions guide the ethical investigation of a technosolution, which I operationalise here as assessing whether its problem construction, sociopolitical dimensions, scope, and legitimacy can withstand scrutiny. Or, to put it differently, whether the approach is genuinely oriented toward addressing a problem and honest about its limitations. In the next section, I use this framework to scrutinise Q4G as the normative ambition to leverage quantum technologies for addressing social challenges such as climate change and hunger.

Overall, this framework responds to the need articulated by Siffels and Sharon (2024) to develop ethical interventions *ex ante*, ensuring that ethical scrutiny occurs before problematic technosolutionism becomes entrenched in the justification, design, and deployment of a technology. These guiding questions offer a means to critically assess the notion of Q4G in its embryonic stages, particularly the search for quantum use cases, at a stage when there is still room for genuine debate about the ethical and sociopolitical dimensions of such proposed technosolutions.

5. Assessing Quantum for Good through a lens of technosolutionism

Building on the preceding discussion of technosolutionism, Q4G is assessed both as a normative ambition and with respect to how it has been operationalised through the search for use cases. The assessment follows the guiding questions outlined in the previous section regarding problem construction (5.1), sociopolitical conditions (5.2), scope (5.3), and legitimacy (5.4).

Throughout this analysis, I use the Open Quantum Institute (OQI) as a concrete example of a Q4G initiative, dedicated to "the development of applications for the benefit of humanity" (Open Quantum Institute, n. d.). As the most systematic effort to date to advance the Q4G agenda, the OQI offers a valuable point of reference for our analysis. The analysis draws on the use cases that the OQI has formulated to mobilise quantum technologies in support of the SDGs, using these as illustrative examples of envisaged quantum technosolutions.

Table 1 summarises the use cases formulated by the OQI. Both the selected SDG-related problems and their proposed quantum approaches are taken directly from the OQI White Paper (Open Quantum Institute, 2024). Notably, OQI's own readiness rankings indicate that the vast majority of these proposed applications remain at an early, conceptual stage, mostly expressed as initial ideas or high-level proposals with little technological maturity at present. The fifth column highlights aspects of the proposed quantum solutions that may raise ethical concerns. I focus primarily on ethical concerns directly associated with the specific use case, including those arising from exacerbating existing inequalities or the inheritance of risks associated with other

technologies (e.g. algorithmic bias). Some of these concerns are reminiscent of issues arising around other new and emerging technologies. That does not, however, make them less deserving of our attention in the context of quantum technologies; on the contrary, as has been argued elsewhere (Shelley-Egan & de Jong, 2025), attending to these continuities – rather than fixating on what is novel – enables us to draw on analogous discourses and avoid reinventing the wheel (see also Swierstra & Rip 2007).

Beyond the considerations raised by the specific use cases, I acknowledge broader ethical issues that cut across all proposed use cases, particularly (global) accessibility and dual-use risks. Given the high cost and resource-intensive nature of current quantum technologies, accessibility constitutes a systemic ethical concern that extends beyond any individual use case. Likewise, the (aspirational) technological capabilities on which these use cases rely carry inherent dual-use potential, with possible beneficial and harmful applications.

While these general concerns are important, they sit slightly apart from the aim of evaluating the proposed use cases as technosolutions. This analysis therefore keeps these cross-cutting issues in view, but focuses on the ethical questions that arise directly from positioning quantum technologies as solutions to SDG-related problems.

Furthermore, I do not examine each use case in detail; instead, I use them to illustrate key issues in the broader Q4G frame. What follows should thus not be read as a comprehensive ethical assessment of the OQI's individual use cases, but as a critical examination of the overarching Q4G frame through the guiding questions developed above, identifying aspects of its technosolutionist logic that are ethically questionable.

Table 1 gives an overview of OQI's quantum use cases and the ethical concerns they raise. While not every case is discussed explicitly, the table shows the range of technosolutions in OQI's portfolio and their potential ethical issues. Section 5.1 selectively draws on some of these cases to illustrate the ethical assessment of Q4G.

Table 1. OQI's quantum use cases: a selection of SDG-related problems and their proposed quantum solutions as presented in the OQI's White Paper (2024)
(table continues on the following pages)

SDG	*Theme*	*Problem*	*Quantum solution*	*Ethical concerns*
2 – Zero Hunger	Plant Genomics	Unlocking the genetic basis of complex plant traits for sustainable agriculture is hampered by the	Quantum computing offers the potential to overcome these bottlenecks by accelerating genome-wide association	*Concentration of power:* Quantum-accelerated gene editing may reinforce existing monopolies in agribusiness by giving

Table 1. (continued)

SDG	*Theme*	*Problem*	*Quantum solution*	*Ethical concerns*
2 – Zero Hunger *(continued)*	Plant Genomics *(continued)*	massive size of crop genomes and the computational limitations of current methods for analysing high-dimensional genomic data.	studies and simulation of complex gene networks, thereby supporting more efficient and scalable gene editing for improved crop performance.	corporations unprecedented control over crop genetics and proprietary genomic data, potentially marginalising smallholder farmers, restricting genetic diversity, and deepening global agricultural inequality. *Ethical acceptability of gene editing:* The use of quantum computing to model and accelerate gene editing raises questions over the naturalness and moral legitimacy of modifying plant genomes at scale.
	Last-Mile Food Delivery	Inefficient last-mile food delivery systems contribute to significant food waste, high costs, and unequal access. Traditional methods struggle with the computational intractability of optimally solving complex problems like the Vehicle Routing Problem (VRP).	Quantum computing, particularly through hybrid quantum-classical approaches, could optimise formulations of the Vehicle Routing Problem (VRP), could enable more efficient routes and help lower the accessibility and affordability of food distribution systems.	*Equity of benefit distribution:* There is a risk that quantum-optimised logistics will primarily benefit large corporations, leaving smallholders and low-income communities underserved, or excluded. *Environmental impact:* While aiming to reduce food waste, quantum-optimised routing could increase food production to meet new demand elsewhere (e.g. through rebound effects) or increase packaging, offering sustainability gains.

Table 1. (continued)

SDG	*Theme*	*Problem*	*Quantum solution*	*Ethical concerns*
3 – Good Health and WellBeing	Accelerating Novel Antibacterial Discovery	The discovery of novel antibiotics is hindered by limited predictive tools for understanding molecular interactions and drug resistance accumulation, making the exploration of vast chemical space slow, expensive, and too risky to compete with other drug types.	Quantum reservoir computing and quantum simulation offer a promising approach to modelling complex molecular interactions, enabling faster, more efficient resistance prediction and improved integration of large-scale drug information to support translational research.	*Environmental impact:* Accelerated drug development could increase pharmaceutical waste or ecological disruption if not governed. *Alignment with public health priorities:* Quantum drug design may prioritise profitable compounds over urgent global health needs, potentially neglecting equitable access or overlooking antimicrobial resistance challenges.
	Predicting Gastrointestinal Cancer	The early detection of gastrointestinal cancers is hindered by the sheer volume of data generated by non-invasive diagnostic technologies, overwhelming both human and clinical computational resources.	Quantum machine learning approaches – such as quantum neural networks and quantum support vector machines – could help address these bottlenecks in real-time diagnosis and earlier intervention for colorectal cancer.	*Fairness in clinical AI:* Automated quantum diagnostics might reduce reliance on clinical judgment or human oversight.
6 – Clean Water and Sanitation	Water Leak Detection	Urban water systems suffer from significant water loss due to leaks, but optimising a limited number of sensors to detect those leaks is hard, and current methods address it inefficiently.	Quantum optimisation techniques, particularly quantum annealing (such as QUBO formulations), could efficiently solve the sensor placement problem on large, complex water	*Dependence on technology:* Over-reliance on complex, community-led water systems may reduce resilience in simpler, community-led leak detection and repair. Digital infrastructure could be vulnerable to hacking or misuse,

Table 1. (continued)

SDG	*Theme*	*Problem*	*Quantum solution*	*Ethical concerns*
6 – Clean Water and Sanitation *(continued)*	Water Leak Detection *(continued)*		networks, enabling better leak detection and reduced water loss.	posing risks to water supply reliability and public safety.
	Eliminating 'Forever Chemicals' From Water Sources	PFAS ('forever chemicals') are highly persistent toxic contaminants in water that are extremely difficult to detect and remove, posing severe health and environmental risks to the public.	Quantum simulation methods can accurately model the complex chemical bonds in PFAS molecules, enabling a more efficient design of catalysts and treatments to destroy these contaminants and improve water treatment technologies.	*Environmental safety:* New materials or processes designed to break down PFAS might generate harmful by-products or have unforeseen ecological impacts if not thoroughly tested.
	Molecular Docking for Chemical Pollution	Phenol and other persistent organic pollutants in water bodies pose significant health and environmental risks, yet current degradation methods are energy-intensive, costly, or slow, with limited effectiveness.	Quantum simulation, combined with machine learning at a molecular level, enables the design of more efficient bioremediation catalysts and treatment strategies.	*Social and economic displacement:* More efficient industrial cleanup technologies might accelerate the replacement of workers related to traditional remediation methods, affecting livelihoods without adequate retraining programmes.
7 – Affordable and Clean Energy	Layout of Turbines in a Wind Farm	Optimising wind farm turbine layouts to maximise energy output is a complex, combinatorial problem due to wake effects and large-scale variables	Quantum optimisation methods offer a promising approach to solving wake interaction configurations, enhancing wind farm efficiency and	*Environmental impact:* Algorithmic decisions may overlook local ecological or social consequences.

Table 1. (continued)

SDG	*Theme*	*Problem*	*Quantum solution*	*Ethical concerns*
7 – Affordable and Clean Energy *(continued)*	Layout of Turbines in a Wind Farm *(continued)*	limiting the efficiency and reliability of renewable energy potential.	supporting the transition to clean, affordable energy.	
	Smart Grid Management	Managing large-scale electrical grids efficiently requires balancing the production, integration, and consumption of energy in real-time, involving the complexity of uncertain renewable energy sources, the need for optimisation, integration of intermittent renewable energy resources, and improved management of transmission losses, emissions, and stability.	Quantum optimisation algorithms like QAOA offer promising approaches to complex grid management problems, enabling faster, more efficient energy dispatch and improved management of energy distribution, transition energy systems.	*Applicability challenges in low-income countries:* Implementation may be constrained by structural factors, such as limited electrical infrastructure, expertise, and resources, potentially hindering equitable deployment. *Security risks of quantum-controlled infrastructure:* Reliance on quantum-controlled systems for critical systems like grid management raises concerns about vulnerabilities to cyber threats or misuse, threatening security.
13 – Climate Action	Catalytic Carbon Capture	Modelling and optimising catalytic materials for carbon capture is computationally demanding due to the complexity of molecular interactions and number of possible reactions, hindering the development of efficient, scalable solutions to reduce CO_2 emissions.	Quantum machine learning methods can overcome classical limitations by efficiently exploring complex chemical spaces, enabling the discovery of optimal catalytic materials that improve carbon capture performance and accelerate climate action.	*Environmental trade-offs:* Focus on carbon capture may divert attention and resources from emissions reduction or social justice in climate policy.

5.1 Problem construction

5.1.1 A solution looking for problems

What, precisely, is the 'good' in Q4G? The term typically functions as an umbrella concept for quantum technology applications presumed to serve the public good. The OQI defines the 'good' of Q4G through alignment with the SDGs, translating broad societal aims – such as health, food security, and climate action – into technically addressable sub-problems (e.g. drug optimisation, supply-chain logistics). This can be understood as an attempt to channel quantum innovation towards socially desirable goals and thus resonates with responsible innovation ambitions.

The underlying assumption here is that quantum technologies will be beneficial. This echoes a broader solutionist logic in which emerging technologies are treated as inherently valuable, even before feasibility or effectiveness is demonstrated. This combines a speculative and a largely backward-looking orientation: it begins with quantum's perceived potential and works backwards to identify problems that can be reframed to match. So, although Q4G appears to start from real-world challenges (e.g. the SDGs), the technological means appear predetermined. And importantly, these means are still speculative (see also Section 5.3).

Such a 'technology-first' approach – starting from a technology and looking for problems it might address – is not inherently problematic, though it has two effects that both carry risks: shaping which problems receive attention, and shaping how those problems are defined. The first is discussed below; the second in the next section.

The first risk arises because the existence of quantum technologies, and the ambition to demonstrate their societal value, shape which problems receive attention. I refer to this as an *agenda-setting effect*. Within the space of SDG-relevant challenges, those that can be framed in quantum-amenable terms, such as optimisation tasks and simulation problems, are more likely to be selected than those that cannot, regardless of whether they represent the most urgent or impactful entry points. Again, this need not be problematic in itself. The concern arises when the capability-driven character of this selection goes unacknowledged. If a technology-first approach shapes which problems get selected, there is a risk that technological opportunity – rather than social urgency – becomes the leading principle for allocating attention and resources to global challenges, with resources flowing toward those aspects of a problem for which a technological solution exists rather than those that are most pressing.

This risk is sharpened when the social legitimacy of the SDGs is invoked to validate an innovation agenda whose actual contribution to those goals remains uncertain or marginal. The SDGs carry significant normative weight: they represent internationally recognised commitments to address grand societal problems such as hunger, disease, and climate change. Associating quantum innovation with them under the banner of

'for good' borrows that legitimacy. If the contributions being claimed are highly speculative, or unlikely to contribute significantly to the societal goals or to reach those most affected, that association risks functioning as 'legitimacy washing': instrumentalising the legitimacy of the SDGs to advance an independent innovation agenda. The risk here lies in (unintentionally) creating the impression that more progress can be made on urgent challenges than can reasonably be expected – in particular when the technology at hand is highly speculative – and obscuring the gap between who the SDGs are meant to serve and who is likely to benefit from quantum innovation in practice.

To avoid the tacit instrumentalisation of the SDGs, quantum contributions to SDG goals should therefore be presented with transparency about their speculative character and the distance between what is technologically feasible and what is socially needed; Section 5.3 further discusses this issue of scope. This agenda-setting effect is closely related to a second concern: a technology-first approach shapes not only which problems receive attention, but how those problems come to be defined. I turn to this next.

5.1.2 Depoliticising problems

The second risk concerns how problems are defined once selected. I call this *problem narrowing*: the risk that complex societal challenges are redefined in terms of what quantum can address, excluding dimensions of the problem that are morally significant but not reducible to technical terms. When hunger is reduced to supply-chain logistics, or climate change to carbon capture optimisation, something normatively significant is lost. The structural dimensions of these challenges – i.e. the social and political aspects that are often central to their persistence – risk to be backgrounded when the problem is defined around technological fit. In other words, the problem gets *depoliticised*: stripped of the context that would be essential for addressing it meaningfully. This is not a fault of the OQI's mandate, which is explicitly technocentric. It does, however, highlight the limits of what a technosolutionist framing can achieve for systemic challenges like the SDGs. To avoid overlooking the social, economic, and political dimensions of these issues, it is important to recognise the limits of technosolutions and position them as one tool within a broader, multi-faceted effort to advance the SDGs.

Problem narrowing also has a further consequence: once a problem has been defined around its technically tractable dimensions, solutions that address the broader problem naturally fall out of view. Alternatives that may be more effective or ethically preferable, including non-technological interventions or approaches that address root causes rather than symptoms, risk being sidelined simply because the problem definition overlooks them, not because they have been evaluated and found inferior. This problem of overlooked alternatives taken up further in Section 5.4.

5.2 Sociopolitical dimensions

5.2.1 Motivation: 'good' for whom?

The motivations behind Q4G are normatively framed: to benefit society, address global challenges, and support the public good. This also holds for the OQI's initiative to leverage quantum technologies for the SDGs. In this way Q4G is being positioned as a socially responsible technology initiative, or simply: a form of responsible innovation. However, these aspirations 'for good' remain under-scrutinised: whose good, by which values, and at what cost?

Engaging with these questions is essential for evaluating both the urgency and the legitimacy of proposed quantum interventions. Yet within the current Q4G discourse, these questions are rarely addressed in a systematic way across use cases and initiatives. As a result, the label 'for good' risks functioning more as a normative veneer than a genuine, substantive ethical commitment. Without critical engagement with what 'good' means here, Q4G risks becoming little more than an ethics-washed version of 'quantum for profit'.

Take the OQI's use case of last-mile food delivery which aims to use quantum optimisation to reduce inefficiencies in food distribution systems. The envisioned outcome is the minimisation of waste and delivery costs, thereby contributing to SDG 2: Zero Hunger. But if this use case is implemented in settings where food scarcity is not a pressing issue, or if the primary beneficiaries are commercial logistics providers in high-income countries, then we must ask: Which hunger problem is being addressed, and who actually benefits? Are those who stand to benefit from quantum innovation also those the SDGs are most concerned with? Given that quantum development is currently concentrated in wealthy institutions and high-income countries, the benefits of quantum innovation are likely to materialise there first, thereby potentially widening rather than narrowing existing global inequalities.

This example shows how a technical problem – optimising distribution – can be framed as contributing to a social goal, even when its real-world impact on that goal remains unclear or marginal. It underscores the risk of technological aspirations, rather than ethical urgency or social need, driving the selection of use cases, blurring what the notion of 'for good' actually entails.

5.2.2 Criteria for success: technical progress as social value

Because Q4G is loosely defined, the criteria for judging its success are similarly open-ended. While initiatives such as those of the OQI link quantum innovation to the SDGs, progress is often described in terms of technical achievement or application potential, rather than demonstrated benefit for people or communities. This reflects

an assumption that if quantum tools can be applied to domains such as drug discovery, logistics, or climate modelling, social value will naturally follow.

To ensure that technical achievements translate into genuine public benefit, however, assessments of viability, scalability, and desirability in real-world social contexts are essential. Equally, the process of setting priorities and defining success needs to involve those most affected by the challenges Q4G seeks to address. At present, decisions appear concentrated among institutional actors and technology developers, reinforcing a technocratic model in which technical advancement is equated with progress. In complex social domains such as hunger, health, and climate, success cannot be inferred from what is technically possible, but should be negotiated and socially determined.

The OQI's antimicrobial resistance (AMR) use case illustrates the limitations of purely technical approaches to global challenges. It proposes using quantum machine learning to accelerate antibiotic discovery, a goal aligned with SDG 3: Good Health and Wellbeing. The proposed solution centres on expanding the regions of chemical space that can be simulated, improving in silico predictions, and enhancing molecular modelling. While these are valuable scientific aims, they do not engage with the social and political factors that ultimately determine whether new antibiotics are equitably accessible and effectively used, which is key to genuinely addressing SDG 3.

This is not a criticism of the OQI's approach, which is rightly technical. It does, however, highlight a broader point: if quantum technologies are to contribute meaningfully to SDG 3, the question of access – particularly in low- and middle-income countries, where AMR burdens are greatest – must also be addressed. The same holds for transparency about the speculative character of the envisioned solutions, which is the focus of the next section.

5.3 Scope

5.3.1 Temporal scope: long trajectories, urgent challenges

Q4G is imagined as a long-term vision, aimed at tackling major societal and global challenges. The search for use cases can be understood as identifying near- or mid-term goals that contribute to these challenges. Such use cases often highlight what quantum technologies might achieve, but their actual social value – especially in the near future – remains uncertain, limited, or still difficult to demonstrate.

The tension here lies in the gap between the urgency of the challenges (e.g. food insecurity, antimicrobial resistance, climate breakdown) and the extended, uncertain timelines of quantum technology development. Designing interventions around capabilities that do not yet exist gives rise to the question of whether quantum technologies can be legitimately linked to such concrete and urgent problems, when their viability has not been proven yet.

As argued in Chapter 5, this is precisely the epistemic situation in which a technology is 'not ready' yet for an ethics oriented towards outcomes: when a technology is still largely speculative, the ethical analysis of its projected outcomes is itself speculative, and risks directing attention to issues that may not turn out to be the relevant ones. Q4G's framing of quantum innovations as contributions to urgent SDG challenges sits in tension with this, as it engages in outcomes-oriented reasoning at a stage where there are still fundamental uncertainties about the development of a technology and its applications.

This tension is illustrated by the OQI's use cases of optimising last-mile logistics and modelling in plant genomics. These innovations might contribute to SDG 2: Zero Hunger, but they might not address the most pressing forms of hunger or crop vulnerability in practice. The point here is that if meaningful impact is only expected decades from now, how does this serve problems that require immediate intervention? The honest answer is that it may not serve those problems directly. This gap should be acknowledged rather than obscured by the 'for good' framing. Without clarity about time horizons, there is a risk of overclaiming relevance, or even displacing resources from more immediately effective approaches.

Even modest attempts to align quantum initiatives with the SDGs carry potential risks when they rely on speculative interventions, particularly if public funding, policy support, or state endorsement is involved. Such investments could, in principle, divert attention from more proven or immediately effective responses. The extent to which this occurs in practice remains to be traced and warrants further investigation. Framing these speculative quantum initiatives as 'contributions' despite their uncertain impact can obscure these trade-offs and create a misplaced sense of urgency or appropriateness.[8] Moreover, investment in commercial and security-oriented quantum development may be mobilised to bolster the credibility of Q4G initiatives, creating a legitimising dynamic that persists even in the absence of successful practical applications in those domains.

5.3.2 Impact: techno-optimism outpacing ethical reflection

The presence of techno-optimism in the quantum realm was already signalled in Chapter 6, where I discussed the visions currently driving the development of quantum computing (Chapter 6, Section 4.1). The phenomenon of Q4G is similarly underpinned by optimism about quantum capabilities and their potential for societal benefit. While such optimism can be a productive force, motivating innovation and mobilising resources, it carries ethical risks when it goes unscrutinised.

[8] This tension between promised benefits and uncertainty is a recurring pattern across emerging technologies, and not unique to the quantum domain.

In Chapter 6, I already pointed to the risk of uncurbed optimism hindering responsible decision-making. Optimistic claims may generate a dynamic of hype, attracting investment and policy support without adequately acknowledging the uncertainties involved. Moreover, assuming quantum technologies to be beneficial may tacitly bypass the prior ethical questions of whether and how they should be developed, and for which purposes.

Additionally, prevalent optimism carries the risk of selective attention: by foregrounding potential benefits, risks and limitations are pushed to the background. While the language of *intended* impact is ubiquitous in the Q4G discourse, there is little evidence of systematic engagement with potential *unintended* consequences or downsides of scaling quantum technologies. Issues such as the energy demands of quantum computing, material extraction for hardware, or the carbon footprint of data infrastructure remain largely underexposed (see Table 1). This raises the concern that the techno-optimism that drives the ambition of Q4G is outpacing careful ethical and ecological consideration.

Finally, uncritical optimism may tilt toward *distributional blindness*: by focusing on aggregate benefits, the distributions of those benefits remain invisible. As discussed in Section 5.2, the benefits of quantum innovation are likely to materialise first in high-income countries and wealthy institutions. An optimism that does not reckon with this risks reinforcing existing inequalities under the banner of 'for good'.

5.4 Legitimacy

5.4.1 Fit for purpose

Q4G rarely claims that quantum is the only solution to societal challenges, yet it often presents quantum technologies as uniquely positioned to make significant contributions to addressing them. Without systematically comparing these proposals to non-quantum alternatives, advocacy for quantum solutions risks obscuring approaches that may be more effective or appropriate. For instance, while climate modelling is often cited as a prime Q4G use case, classical computing methods may remain more practical and scalable in the near term – not least given the projected energy demands of quantum computers large enough to perform such modelling (see Table 1). It is often unclear how the value of the proposed quantum approaches compares to non-quantum alternatives.

This points to a central legitimacy question, focused on comparative effectiveness: Is quantum truly the best technological fit for a given problem, or is it being advanced primarily because it aligns with strategic priorities or funding opportunities? In the OQI's use cases, this question tends to be bypassed in favour of finding a fit for quantum rather than critically evaluating whether quantum offers the most

appropriate response. This tendency directly connects to the problem construction concerns discussed in Section 5. Genuinely assessing fit for purpose would require comparing quantum solutions to non-quantum alternatives on relevant dimensions such as effectiveness, scalability, cost, accessibility, and timeline to impact. Notably, this comparative assessment is not primarily a task for initiatives like OQI, whose mandate is to develop and demonstrate quantum applications. The responsibility lies rather with those who decide whether and how to support and endorse specific such approaches, such as research funders and policymakers.

5.4.2 Technofix or system fix?

Prior to the question about fit for purpose comes a more fundamental question: To what extent can any technological solution address challenges as socially embedded and structurally complex as hunger, climate change, or global health? As discussed in Section 5.2, Q4G's focus on use cases tends to cast global challenges as technical problems to be solved, rather than complex systems shaped by structural inequalities, political choices, and power. Effectively addressing issues such as climate change and global health is not a mere design problem; it requires systemic change of how we live our lives and organise our societies.

By focusing on technological mitigation of symptoms many Q4G proposals risk obscuring the need for broader interventions to tackle root causes. For example, while quantum-enhanced carbon capture could aid climate mitigation, emphasising this technological solution could draw attention and resources away from other systemic measures – such as emission reductions and climate justice initiatives – that are essential for genuinely addressing climate change. Similarly, quantum-enabled antimicrobial discovery could accelerate drug development, but without accompanying efforts to address global health inequities, its contribution to improving global health (SDG 3) will remain limited.

This is not a critique of Q4G as such, but of how its role is positioned and legitimised. A more responsible framing would acknowledge that quantum solutions address one dimension of a complex challenge, while explicitly supporting – rather than displacing – complementary efforts to address structural dimensions. Whether quantum technologies are appropriate or desirable solutions is not a question that can be answered in general terms. This must be assessed case by case, in relation to specific contexts. For Q4G to constitute a form of responsible innovation, it must invite this kind of critical scrutiny rather than assume its own relevance by default.

6. Conclusion

Quantum for Good (Q4G) has recently emerged as a normative ambition to harness quantum technologies for the public good. Initiatives under this banner aim to align

quantum innovation with societal challenges, most notably by identifying use cases promising social benefit.

This approach – seeking to address social problems with technological means – bears the hallmarks of technosolutionism. In this chapter, I used technosolutionism as an analytical lens to critically yet constructively examine Q4G. I adopted a nuanced perspective, treating technosolutionism as something that can be but is not inherently problematic. Instead, I argued that its ethical status in any specific case depends on how the problem is constructed, what motivations underlie the proposed solution, and whether the intervention can be considered ethically appropriate. To investigate this, I formulated a set of guiding questions about the way a problem is constructed, how its sociopolitical dimensions are addressed (or sidelined), and how the scope and legitimacy of the solution are framed. I used these questions to evaluate the ethical dimensions of the technosolutions promoted under the banner of Q4G.

Through its normative aspiration, Q4G gestures towards more responsible and societally attuned technological innovation. Linking quantum research and development to societal goals – such as the UN Sustainable Development Goals (SDGs) – marks a welcome shift from purely military or commercial imperatives towards more socially engaged innovation, actively steering quantum innovation in the public interest. However, analysing the quantum use cases formulated by the Open Quantum Institute in service of the SDGs, this chapter identified several pitfalls associated with the technosolutionist logic of Q4G.

The technology-first orientation of Q4G bases the selection of problems on technological capability rather than social urgency. When done under the banner of 'for good', the legitimacy of the SDGs risks being instrumentalised to establish an innovation agenda whose actual contribution to those goals remains uncertain. Additionally, it narrows issues to technologically tractable problems. Complex societal challenges are 'technologised' or 'depoliticised': they are redefined in quantum-amenable terms, backgrounding their structural and political dimensions.

Second, the 'for good' in Q4G remains ambiguous as long as questions of whose good, by which values, and at what cost go unaddressed. Currently, the innovation ecosystem promoted under Q4G is not clearly separable from commercial or strategic agendas. In the absence of transparent and socially grounded success criteria, Q4G risks becoming a normative veneer: a legitimising narrative for 'quantum for profit' or 'quantum for competitiveness'. There is nothing inherently wrong with pursuing economic or strategic goals, but these should not be conflated with the pursuit of the public good, and not taken to lend credibility when practical applications in these domains remain absent. For Q4G to go beyond rhetorical alignment with public good (e.g. the SDGs), it must engage with the ethical and sociopolitical conditions for real-world impact, and not treat these as secondary issues.

Third, given the early-stage nature of most quantum technologies and the uncertainties surrounding their development, claims about their value for urgent societal issues often rest on speculative assumptions. This raises questions about whether quantum technologies can realistically be considered a plausible and appropriate solution. As argued in Chapter 5, this speculative character makes it epistemically risky to make ethical claims about the societal benefits (or risks) of quantum technologies at this stage. Q4G's framing of quantum innovations as contributions to urgent societal challenges sits in tension with this: it engages in outcomes-oriented reasoning before the technology is actually ready for such engagement. At a minimum, the discourse on Q4G should be more transparent about the uncertainties involved and the distance between what is currently feasible and what is socially needed.

Finally, the legitimacy of technosolutions requires 'fit for purpose'. Assessing this means asking whether quantum solutions compare favourably to alternatives on dimensions such as effectiveness, scalability, cost, accessibility, and timeline to impact. Given the early stage of quantum development and the speculative character of many of the use cases examined here, their fit for purpose is contestable on the dimension of timeline alone – and this is before asking the more fundamental question of whether a technological approach is the right kind of response to challenges as structurally complex as hunger, climate change, or global health.

For Q4G to fulfil its promise and avoid becoming a mere legitimising narrative for existing innovation agendas, it must honestly engage with the limits of technological fixes and be transparent about uncertainties rather than being carried along by highly speculative scenarios. Only then can Q4G move from symbolic ambition, or worse, a normative façade, to public benefit. Here lies an important task for the research community, particularly for researchers focusing on the ethical, legal, social aspects of technology. However, critical scrutiny of technosolutionist initiatives extends beyond academia; the guidelines outlined in this chapter may also be integrated into broader technology assessment approaches.

To conclude, the aim here has not been to dismiss Q4G for its technosolutionist tendencies. Rather, by drawing attention to these tendencies, this chapter highlights the risks embedded in current framings and practices of Q4G. Framing Q4G as a form of technosolutionism is intended not as rejection but as an analytic tool to invite critical reflection within the field. Such scrutiny supports the development of a more ethically robust paradigm. The guiding questions proposed here offer a means to identify ethical issues early in the design of technological interventions – quantum or otherwise – and to address morally problematic forms of technosolutionism before they become entrenched.

Chapter 8

Risk and Precaution in the Quantum Discourse

An earlier version of this chapter is under review as: de Jong, E., & De Haro, S. (2026). Better secure than sorry? Precaution in the quantum-safe cryptography transition. In: Ethics and Information Technology.

Abstract

Full-fledged quantum computers are expected to be able to break the public-key cryptographic systems securing most digital communications, financial transactions, and state secrets. Although no such machine exists today, a discourse has emerged that strongly calls for a society-wide transition to quantum-safe cryptography in anticipation of this threat. Such a transition demands substantial investment of time, resources, and organisational effort in response to a risk whose inducing cause – a cryptographically relevant quantum computer – remains deeply uncertain. In much of the policy and expert discourse, however, the conditional 'if' of the arrival of such a machine has been tacitly replaced by 'when', thereby sidelining the profound uncertainties involved. As argued in Chapter 5, ethical engagement with the potential outcomes of a still-speculative technology demands explicit justification. This chapter holds the anticipatory discourse surrounding the quantum cyber security threat to that standard, asking: When is it ethically defensible to treat an uncertain technological development as a risk that warrants precautionary action? The chapter develops a philosophical framework for assessing the ethical defensibility of risk claims and precautionary responses. Drawing on the notion of technological visions, it argues that risk claims and precautionary proposals about still-speculative technologies are best approached as visions – normatively charged ideas about a future technology that motivate present action – and analysed across three levels: their substantive content, the argumentative moves that make them normatively compelling, and the institutional actors who advance them. Applying the resulting framework to the quantum cyber security discourse makes the normative reasoning in this debate explicit. This adds to a broader understanding of the cyber security threat posed by large-scale quantum computers, while also revealing where that understanding remains limited. The analysis supports the dominant narrative: the threat appears plausible and the QSC transition proportionate. However, it also reveals that this assessment depends on assumptions that usually remain largely implicit. The chapter argues that making these assumptions visible is a condition for ethical defensibility, and – by enabling them to be contested – a prerequisite for democratic debate about the governance of emerging technologies.

1. Introduction

Among second-generation quantum technologies, significant attention goes to quantum computing. One important reason is the concrete risk associated with it: sufficiently powerful quantum computers are expected to break the public-key cryptographic systems that currently secure most digital communications, financial transactions, and state secrets. Large-scale quantum computing is thus widely seen as an unprecedented, society-wide cyber security threat. For societies as digitally dependent as ours, this prospect confronts us with a serious problem.

The good news is that no such machine exists today. Expert estimates of when one might be built vary widely, but even optimistic projections place it at least a decade away (Mosca & Piani, 2026; Doyle et al., 2026). Yet this reassurance may be deceptive, for transitioning to cyber security protocols that can withstand quantum attacks is complex and will take many years (Joseph et al., 2022; Kong et al., 2022). Given the long lifespan of sensitive data, we would effectively have had to start yesterday securing today's data against tomorrow's adversaries. With the research and development of quantum computing steadily progressing, the urge to anticipate a quantum computer that is capable of breaking modern cryptography – in short: a cryptographically relevant quantum computer (CRQC) – grows.

The hypothetical moment when such a machine becomes capable of breaking current encryption – often referred to as 'Q-day' – has therefore become a focal point of strategic and ethical concern. In response, policymakers, industry actors, and researchers increasingly call for a rapid transition to *quantum-safe cryptography*[1] (QSC): encryption methods designed to resist quantum attacks. After years of standardisation efforts, such technically viable schemes are now available (National Institute of Standards and Technology, 2024). At first glance, then, the situation appears manageable: a foreseeable risk that can be met with a ready technical solution.

However, while the transition to QSC has repeatedly been framed as a prudent precautionary – to some extent even a no-regret – measure, the complexity of the operation is not to be underestimated. It is far from a routine software update: public-key cryptography is deeply embedded in digital infrastructures, and replacing it requires extensive re-engineering, coordination, funding, and time. In this case, precautionary action is therefore costly and demanding.

Whether such investments are a wise use of resources for those who bear their costs is itself open to debate. The quantum case might, in this respect, appear typical for

[1] Quantum-safe cryptography encompasses two main strategies to protect against quantum attacks: *post-quantum cryptography*, which adapts classical algorithms to remain secure against quantum computers, *and quantum cryptography*, which uses quantum-mechanical principles – such as quantum key distribution – to achieve quantum-resistant encryption.

classical risk analysis. Yet beyond competing assessments of prudence lies a more fundamental philosophical issue. The question is whether it is ethically defensible to impose tangible and immediate costs in response to a risk whose cause remains speculative. I take a course of action to be ethically defensible when it can be justified through explicit, transparent reasoning that engages seriously with the relevant ethical considerations, even if reasonable people might weigh those considerations differently and reach different conclusions. It is thus primarily a procedural standard for the ethical robustness of reasoning.[2] What is ethically at stake here is the possibility that resources are wasted on a threat that never materialises, and that attention and funding are diverted from more urgent and certain priorities.

The precautionary reasoning in this discourse rests on a logic of *if…then…so…*: *if* a quantum computer capable of breaking current encryption is developed, *then* cyber security is unacceptably compromised, *so* precautionary action is warranted. In much of the policy and expert discourse, however, the 'if' tends to be tacitly replaced by a 'when'[3], as though the arrival of a CRQC were only a matter of time (Mehnert 2025, p. 6; see also Chapter 6)[4]. Illustratively, the Quantum Threat Timeline Report 2025 (Mosca & Piani, 2026) notes: "No fundamental barrier to creating a large quantum computer is known, and such an achievement should be considered a matter of 'when' and 'how' rather than of 'if'." (p. 10) This framing treats the eventual arrival of a CRQC as the default assumption. Uncertainty, on this view, concerns only when and how it will happen – not whether it will happen at all.

When the emerging ethical discourse on quantum computing takes this framing as its starting point, it inherits this presupposition rather than subjecting it to independent scrutiny. The result tends to reflect what Nordmann (2007) has described – and criticised – as *speculative ethics*: imagined futures generating moral imperatives in the present. While anticipating risks may indeed express moral foresight, Nordmann's concern is that such reasoning can authorise far-reaching action on epistemically fragile grounds. This is ethically salient because precaution is never costless. It mobilises attention and resources, and may redirect societal priorities in lasting ways.

An asymmetry thus emerges between the speculative cause of risk and the tangible costs of precaution. The QSC transition readily demands immense resources, in terms of e.g. coordination, funding, time. Yet the very event that it anticipates – the arrival

[2] This means an action can be ethically defensible without being the uniquely correct choice, and conversely, an action that happens to produce good outcomes can still fail to be ethically defensible if the reasoning that led to it was opaque, unreflective, or bypassed important considerations.

[3] The Australian Minister for Industry and Science illustratively states about the development of quantum technologies that "it's no longer 'if', but 'when'" (Australian Government, 2023).

[4] Chapter 6, which analyses the visions of quantum computing, identifies a pervasive techno-optimism when it comes to the development of quantum computing, portraying progress as steady and linear and its arrival as inevitable.

of a CRQC – remains profoundly uncertain. This asymmetry raises an ethical question: How can the imposition of real, immediate costs on individuals and organisations be justified by appeal to a threat whose materialisation remains genuinely uncertain?

According to the ethics readiness framework developed in Chapter 5, ethical engagement with the outcomes of a still-speculative technology is premature and not without its own risks. The quantum cyber security case, however, challenges that default: the potential severity of large-scale cryptographic breaches may justify anticipating concrete consequences despite profound uncertainty. Chapter 5 left this tension open. This chapter takes up that thread, asking under what conditions such engagement may be ethically justifiable. More specifically, it asks: When is it ethically defensible to treat an uncertain technological development as a risk that warrants precautionary action?

To be clear, the ethical issue at stake here does not only concern the financial costs of the transition to QSC. The transition demands broad effort: organisations will need to acquire or develop specialised cryptographic expertise and procure or implement new cryptographic solutions; regulatory frameworks and sectoral guidelines will need to be designed and issued; and the transition will need to be coordinated across deeply interconnected public and private infrastructures and made compatible with other ongoing digital transformations, such as those driven by generative AI (Boulet et al., 2021; Pupillo et al., 2023, p. 7; Capraro et al., 2024; Le et al., 2025, p. 21). In short, this is not a routine technical fix but a large-scale, cross-sectoral undertaking. Precisely because of this breadth and complexity, the precautionary measures it motivates call for robust ethical justification.

This question is largely sidestepped in current calls for a transition to QSC and surprisingly absent in ethical and public debate. Even if the QSC transition can ultimately be solidly justified, the fact that there is no real discussion about *whether* this transition is an appropriate measure, but mainly on *how* to realise it in time, is a risk in itself. Where normative choices go unquestioned, they are made by default by a relatively narrow group of technical and security actors, without being subject to scrutiny or accountability. This contributes to the framing of the quantum cyber security threat as primarily a technical and security matter to be resolved by technical and security experts, when in fact it also involves normative trade-offs about resource allocation, risk tolerance, and societal priorities that go beyond technical expertise. What is missing is an explicit reflection on weighing precautionary reasoning against the costs and trade-offs of the measures it motivates. This chapter seeks to address that gap and investigates the ethical basis of precautionary action in the case of the quantum cyber security threat.

Even if the outcome of such an analysis ultimately ends up giving epistemic justification for the society-wide QSC transition, it will do so in ways that are

illuminating. Ethical analysis is not limited to prohibiting certain actions, but can play a constructive role: by articulating and justifying our reasons for a certain course of action, it can also strengthen the current calls for transition and it can play an important role in securing public acceptance. In particular, the discussion addresses three aspects that such an analysis may cast further light on:

(i) how to weigh the different societal aspects of that transition. In turn, this can help to
(ii) rebalance and fine-tune timelines and plans. Even if the analysis ends up endorsing measures that are currently being taken, it can still help to
(ii) distinguish principled from more opportunistic appeals to precaution.

In later sections, I will return to how the analysis in this chapter contributes to each of these questions.

To set the scene for the philosophical analysis, Section 2 outlines the quantum cyber security threat, its proposed solutions and the practical challenges of transitioning to QSC. Section 3 develops a conceptual framework by examining risk, precaution, and their normative, socially constructed dimensions. Drawing on this discussion, Section 4 develops a framework for assessing the ethical defensibility of risk claims and precautionary measures, approaching them as a special instance of technological visions. In Section 5, this framework is applied to the case of quantum's cyber security threat to assess the ethical defensibility of the currently dominant narrative.

2. The quantum cyber security threat, its solutions and its challenges

2.1 Breaking widely-deployed encryption

In broad terms, quantum technologies harness principles of quantum mechanics to achieve capacities beyond those of technologies based on classical physics. In the case of quantum computing, quantum-mechanical phenomena are used to perform computational tasks. The general ambition is to obtain a computational advantage over classical computers on specific tasks. The advantage over classical computers is not wholesale, but restricted to well-defined classes of problems.

One such problem is factorisation: expressing a number in terms of factors. For example, 2 and 3 are factors of 2, just like 1 and 6, and we can equivalently express 6 as the product of its factors: 2×3 or 1×6. For classical computers it is practically infeasible to solve a special type of factorisation, known as *prime factorisation*:

decomposing a large integer into its prime factors.[5,6] While we commonly like to see our problems solved, mathematical problems are sometimes also used to our advantage. Factorisation – or more specifically: the difficulty of factoring very large numbers into two prime factors – is one of them. Although it is easy to find the prime factorisation of 6, it takes a classical computer a very long time to find the two prime factors of a 600-digit number that was constructed as their product.

The RSA (Rivest-Shamir-Adleman) encryption algorithm exploits this by using numbers that are the product of exactly two large primes, making the factorisation computationally intractable. For a piece of data protected by RSA cryptography, a large integer (a more complex equivalent of the 6 in the example above) which has two prime factors is used as a 'lock' so that no unauthorised parties are able to access the information it protects. Only those who possess the key – i.e. a crucial part of the solution, namely one of the two prime factors (e.g. 3 in the example above) – can solve the problem and access the data. The difficulty of recovering the two large prime factors from their product is the basic principle that keeps our data safe in encryption methods that are widely deployed at present, known as public key encryption.[7]

With the arrival of a large-scale, 'cryptographically relevant'[8] quantum computer (CRQC) this form of encryption becomes seriously compromised. It is expected that

[5] Not every large integer can be factored into two prime factors. For encryption, such integers are first constructed as the product of two large primes and subsequently used as a mathematical problem to encrypt information.

[6] In the example, the number 6 has two possible factorisations: 1×6 and 2×3. However, 2×3 is a special type of factorisation, known as a prime factorisation. This means that each factor is a whole number greater than 1 that has no divisor other than 1 and itself (note that 1 does not count as a prime number). For example, $15 = 3 \times 5$ is a prime factorisation but $12 = 3 \times 4$ is not a prime factorisation because 2 divides 4. For a very large number which has 2 large prime factors, it is practically infeasible for classical computers to determine what the factors are.

[7] The quantum cyber security threat primarily concerns asymmetric encryption schemes such as RSA (Rivest-Shamir-Adleman) and ECC (Elliptic Curve Cryptography), whose security depends on mathematical problems that a large-scale quantum computer could solve efficiently. However, not all cryptographic methods are equally affected. Symmetric encryption, for example, relies on a different type of mathematical structure and is therefore comparatively resilient: although quantum computing would weaken it by making certain attacks faster, the loss of security can generally be offset by using longer keys.

[8] A quantum computer can be considered a threat to currently used encryption methods when it is sufficiently advanced. To break public-key encryption schemes such as RSA, a quantum computer needs to have enough reliable, that is, stable or error-corrected qubits to efficiently run quantum algorithms, e.g. Shor's algorithm. Estimates diverge and highly depend on the assumptions being made about the required quality of the qubits (e.g. significantly more reasonably good or 'noisy' qubits are needed compared to high-quality, fault-tolerant ones), but it is safe to say that a cryptographically relevant quantum computer would require a number of qubits several orders of magnitude more than what current devices possess (Gidney & Ekerå, 2021; Gidney, 2025); although recent research suggests that the number of qubits needed to break for example ECC encryption might be significantly lower than

large-scale quantum computers can perform prime factorisation of large integers efficiently, that is: within a practical timeframe[9], thus effectively breaking the widely-deployed encryption schemes which rely on them. This would allow hostile actors to access encrypted information, interfere with communication and access networks and information systems without permission. It would enable them to view and share information which should remain confidential. This will have far-reaching consequences for digital security across society. Q-day – the projected moment a CRQC becomes operational – has not without reason been dubbed the "cryptocalypse" (Lindsay, 2020).

Secure communication is extremely important in many areas of our lives. Most people use encryption every day, all day: for making bank transfers, securing sensitive information (whether private, business, or state secrets), communicating privately, and signing digital agreements. Encryption is an important cornerstone of a secured digital society and the scenario of currently used methods becoming obsolete *overnight*[10] is alarming. This potential threat is further intensified by the so-called 'harvest now, decrypt later' scenario, in which adversaries collect and store encrypted data today with the intention of decrypting it once CRQCs become available.

2.2 Transitioning to quantum-safe cryptography

Anticipating Q-Day, there has been a near-unanimous chorus of calls to replace current quantum-vulnerable encryption methods with quantum-resistant alternatives. The transition to QSC has become a focal point of strategic and ethical concern. In the United States, the Quantum Computing Cybersecurity Preparedness Act requires federal agencies to prepare for such a transition (Quantum Computing Cybersecurity Preparedness Act, 2022), and the European Union has issued a coordinated roadmap for migration (European Commission, NIS Cooperation Group – PQC Workstream,

previously expected (Babbush et al., 2026; Cain et al., 2026). Mosca and Piani (2026) report a shift in expert opinion towards a higher perceived likelihood that RSA-2048 will be broken, which they attribute to recent progress in quantum computing research and implementations.

[9] In theory, a classical (super)computer could also solve this task, though it would need millions of years to perform it whereas this would take a sufficiently advanced quantum computer a few hours or even less. For example, Mosca and Piani (2026) take a 24-hour period as the relevant timeframe for which the cryptographic scheme RSA-2048 would need to be broken for the risk to be practically relevant. It is worth noting that the difficulty of prime factorisation for classical computers rests on a computational assumption rather than a mathematical proof: no efficient classical algorithm for factoring is known, and centuries of mathematical effort have not produced one, which gives confidence in its hardness. However, the possibility that such an algorithm might eventually be discovered cannot be ruled out entirely. This would represent an additional, distinct threat to current encryption, independent of quantum computing.

[10] The qualifier 'overnight' refers to the theoretical status of current encryption schemes: the existence of a CRQC would render public-key cryptography *mathematically* broken from the moment of its operation. In practice, however, the threat is likely to materialise more gradually, as not everyone will have access to such a device immediately, and attackers will prioritise which systems they target.

2025). After years of consultation, the American standardisation body published a first set of post-quantum cryptographic standards in 2024, providing concrete algorithms for quantum-resistant encryption and digital signatures (NIST, 2024, August 13).

While advocates describe the transition as a prudent precautionary measure – stressing that mapping existing cryptographic dependencies and streamlining fragmented security architectures would be beneficial in any case – many organisations remain hesitant to commit substantial resources in response to what is still a hypothetical threat (Drougkas, 2024; ISACA, 2025). This hesitance may be partly explained by uncertainty: the timeline for the development of a cryptographically relevant quantum computer remains unclear, which may prompt organisations to postpone a complex and costly project.

Public-key cryptography is deeply embedded across digital infrastructures: financial systems, government databases, communication networks, cloud services, industrial control systems, consumer devices, and more. Transitioning may mean replacing a single protocol, but it is more likely to require identifying and updating cryptographic components across many and interdependent layers – many of which are poorly documented, bespoke, and aging which makes the migration effort considerably more complex (Ott & Peikert, 2019; AIVD, CWI & TNO, 2024; Hasan et al., 2024; Loebenberger, 2024; Zelenovic, 2026). Thus this will be a complex and long-term undertaking: judging from previous (smaller) migrations, experts expect that migrating to QSC might take well over 10 years (AIVD, CWI & TNO, 2024). Also, digital infrastructure is typically deeply interconnected, both within and between organisations, involving both public and private actors. This means that a vulnerability in one system can propagate rapidly, and that no single entity can drive the transition alone. Indeed, Robertson et al. (2025) identify coordination as a central challenge in the QSC-transition and a current lack thereof as its central obstacle.

Beyond the technical complexity, transitioning is likely to require substantial investment. To give an indication: the US federal agencies projects that migrating to quantum-safe cryptography will cost the government approximately $7.1 billion (in 2024 dollars) (Executive Office of the President of the United States, July 2024). Moreover, organisations must invest not only in the technical process of replacement, but also in specialised expertise, staff training, software redesign, and in some cases new hardware (Kong et al., 2022).

Beyond cost, the transition may introduce risks and limitations of its own. Post-quantum schemes are comparatively new and have so far been subject to less cryptanalysis – that is, fewer systematic attempts to study or break them – leaving their long-term robustness an open question. They also generally require more computational resources, storage, and bandwidth than the schemes they replace, which may pose challenges in constrained environments. In this sense, the cure is not

guaranteed to be without side effects, potentially introducing new vulnerabilities or performance trade-offs where it seeks to close old ones.

Therefore, transitioning to QSC in anticipation of a future quantum computer capable of breaking current encryption methods will require sustained investments of time, and financial and organisational resources, and is not without risks. It is therefore too simple to frame the migration as a 'better safe than sorry' measure, on the presumption that the cost of acting late necessarily outweighs the cost of acting early. Even if precaution is ultimately warranted, it demands careful ethical justification.

A consideration of these practical costs brings us back to the three questions, (i)–(iii), raised in the introduction, and invites the following reflections: (i) If transitioning to QSC reallocates expertise, public funds and institutional attention, then the proportionality of such investments must be assessed not only in technical terms, but in light of the different societal domains they affect and the alternative priorities they may displace. (ii) This, in turn, bears on how timelines and implementation plans should be calibrated: for example, which sectors warrant accelerated migration, and where a more gradual approach is justified. (iii) Making these trade-offs explicit helps to distinguish principled appeals to precaution – grounded in transparent assumptions about risk and proportionality – from more opportunistic uses of the quantum threat narrative to advance pre-existing institutional or commercial agendas.

As will be argued in Section 3, these practical and distributive questions are central to assessing the 'price of precaution' involved in the QSC transition and to judging when appeals to quantum risk constitute a proportionate, ethically defensible response rather than an opportunistic mobilisation of the threat narrative.

3. Risk and precaution as part of technological visions

In order to investigate what makes it ethically justifiable to treat an uncertain technological development as a risk that warrants precaution, we first need to reflect on the concepts of risk (3.1) and precaution (3.2), which leads into a discussion of their ethical-social dimensions (3.3). I propose the concept of (technological) visions as a fruitful lens to study the claim to risk and the proposed precautionary response in the case of the quantum computing threat to cyber security.

3.1 A philosophical perspective on risk

The term 'risk' is used across a wide range of contexts – from mountaineering and investment to medical treatment and career choices – and the term carries correspondingly varied meanings. Hansson (2023) distinguishes five common definitions: an unwanted event that may occur; the cause of such an event; its probability; its statistical expectation value (probability combined with severity); and

the technical sense used in finance, law, and policy, where 'decision under risk' (known probabilities) is contrasted with 'decision under uncertainty' (unknown probabilities).

While this technical definition dominates many disciplines, the philosophical literature typically uses 'risk' more informally, as the non-trivial probability of an undesirable event (Hansson, 2003). In addition to this characterisation of risk as involving the possibility of future harm (we would not say that there is a 'risk' on benefits), Hansson (2003; 2010) highlights two other aspects salient for the current analysis. First, risk is an *ethically charged* or *value-laden concept*: identifying something as a risk rather than a chance adds a negative evaluation that is not an inherent property of the event itself, but a qualification of it. Second, risk always implies *epistemic uncertainty*: calling something a risk acknowledges that we cannot say for certain whether the unwanted event will occur. This uncertainty can concern the occurrence of a specified event, the severity of the impact and the possibility of unforeseen events, or the cause of the risk itself. I will refer to this respectively as 'occurrence uncertainty', 'impact uncertainty', and 'cause uncertainty'.

Uncertainty, moreover, shapes the bandwidth of interpretation: the less risk judgements can be settled by facts alone, the more room there is for assumptions and interpretive choices to fill it. When probabilities and outcomes are uncertain, deciding whether something counts as a risk – and, by extension, whether it warrants concern – requires normative judgement about the seriousness of potential harms in light of the degree of uncertainty surrounding them.

3.2 The principle of precaution

One response to risk under uncertainty is precaution: acting to prevent or mitigate potential harm even before its likelihood or extent can be fully established. The uncertainty inherent in risk confronts decision-makers with an epistemic challenge when assessing and responding to potential hazards: whether to act on a risk that may not materialise (a costly 'false positive'), or to withhold action on a risk that does (a potentially harmful 'false negative'). Which type of error should be prioritised for avoidance is central to many discussions (and controversies) in risk management. The trade-off between these errors reflects deeper normative questions about what (level of) risks are acceptable and which are not. Acting on the principle of precaution favours the former error (a false positive) over the latter (a false negative).

The principle of precaution (PP) is a norm that guides decision-makers to act preventively in the face of risk. Acknowledging that a single commonly accepted version of the PP remains lacking, Gardiner (2006) suggests that invoking the PP has commonly three components: a threat of harm, uncertainty of impact and causality, and a precautionary response to avoid the harm (p. 6).

At a minimum, this means that 'full certainty is not a precondition for taking precautions' (Steel, 2013, p. 321). In other words, precautionary action can be justified under uncertainty, thereby avoiding decision paralysis under uncertainty.

Gardiner's formulation distinguishes two types of uncertainty that can trigger the PP. The first is uncertainty about the scope or magnitude of the risk. I have called this 'impact uncertainty'. This type of uncertainty particularly applies when its probability cannot be meaningfully specified. We may, for example, put a moratorium on genetically modifying wild animals not because we *know* the risk involved but exactly because we cannot oversee its (harmful) consequences.

The second type of uncertainty concerns 'causality'. While Gardiner does not explicitly define this, he seems to explain this as uncertainty about the link between an event and the materialisation of the harm. In Section 3.1, I called this 'occurrence uncertainty'. There is, however, an additional interpretation of uncertainty about causality: uncertainty not about the occurrence of the impact but of the risk-inducer itself. That is, about whether the activity or technology that would induce the risk will ever materialise. In Section 3.1, I called this 'cause uncertainty'.

In many standard cases of precautionary reasoning this dimension is absent: When assessing the side effects of an existing medical treatment, for instance, there may be uncertainty about whether side effects will occur (occurrence uncertainty) and about their severity (impact uncertainty), but there is no uncertainty about the cause of the risk: the treatment exists and is being administered. Cause uncertainty arises when the risk-inducing cause is itself speculative. This is for example the case when there is uncertainty about the probability of a disease.

This may appear a subtle distinction, but it has direct implications for the justification of precautionary action. Not every conceivable future technology constitutes a risk inducer worthy of precautionary response; that status needs to be 'earned' through explicit justification. The presence of cause uncertainty therefore requires a prior justificatory step: before the costs and proportionality of precautionary action can be assessed, the risk-inducing cause itself must be made sufficiently plausible. The question is not only how to respond to the risk, but whether the speculative cause warrants being treated as a genuine risk-inducer in the first place. This question is, as we will see, pertinent to the quantum cyber security discourse.

Once a risk is deemed sufficiently plausible, the question becomes whether and how the PP can be justifiably mobilised. The conditions for its application remain a much-debated topic (see for example Gardiner, 2006; Manson, 2012; Steel, 2015). One influential approach is Gardiner's 'maximin' approach, modelled onto the Rawlsian maximin principle: choosing the option whose worst possible outcome is least bad – hence 'maximising the minimum' (Gardiner, 2006, p. 13).To avoid the principle's conservative tendency to structurally favour risk-averse choices, Gardiner adds three

criteria[11] for the application of the precautionary principle: 1) significant uncertainty about probabilities or outcomes of the risk, 2) low concern for potential gains (and high concern for protecting a minimum), and 3) alternative options have outcomes that can hardly be accepted. Together, these criteria delineate the scope of cases where it is clear that the PP would demand action (Gardiner, 2006, p. 15).

Gardiner's framework indicates *when* the PP might be mobilised (weighing action against non-action), but says little about what form precaution should take. One instructive starting point for thinking about this is the idea that precautionary action should 'correspond' to the perceived risk (Trouwborst, 2006, p. 150). In other words, the action must be *proportional* to prevent the PP from collapsing into "using a cannonball to kill a mosquito" (Steel, 2015, p. 27).

Steel (2013) develops this requirement by foregrounding proportionality as the key criterion for selecting among possible precautionary measures. In his proposal, proportionality rests on two principles: consistency and efficiency. Consistency requires that precautionary action must not itself cause harms comparable to those it aims to prevent (the cure should not be worse than the disease); and efficiency demands that the chosen measure effectively minimises the target threat while keeping negative side effects as low as possible (pp. 327-328). Steel thus adds an 'unless' clause to Gardiner's framework, specifying that precautionary action is justified unless its negative effects outweigh the harm it seeks to prevent.

While proportionality seems to be a necessary condition for any plausible application of the PP, it does not by itself answer the question of how to weigh the risk of catastrophe against the costs of precaution itself. In other words, what does 'proportional' actually mean in practice? After all, almost no form of precaution is genuinely 'for free': any response comes at a cost, and a coherent notion of precaution must be able to specify how much cost is acceptable in order to enact a given precautionary response.

Munthe (2011) distinguishes three types of cost that together constitute what he calls the "price of precaution": direct costs (resources spent on precautionary measures), indirect costs (risks created by the actions undertaken to manage the original risk), and opportunity costs (benefits foregone by committing resources to precautionary action). Every precautionary decision thus involves a trade-off between these costs and the estimated cost of the risk materialising in the absence of action. For a precautionary response to be ethically justified, this trade-off must be explicitly argued for.

Two aspects thus emerge as central to assessing the ethical justifiability of treating something as a risk and responding to it with precaution: the plausibility of the risk, and the proportionality of the precautionary response. Each involves normative

[11] Modelled after Rawls' own constraints for the maximin principle (Rawls, 1999).

judgement about what type of risk-taking is acceptable. Qualifying something as a risk and defining an appropriate response is, in other words, not simply a matter of *facts* but to an important extent of *values*. Because such judgements are inherently contestable, the question of who gets to make them, on what grounds, and under what conditions becomes itself an ethical question. The next section examines this ethical-social dimension of risk and precaution in more detail.

3.3 Ethical-social dimension of risk and precaution: a vision approach

The previous discussions show that qualifying something as a risk and assessing a precautionary response as appropriate (i.e. proportional) involves normative judgements at multiple levels: in treating an outcome as bad, in judging the harm to be sufficiently plausible, and in deciding both the need for action and its justifiability. These normative judgements together constitute the ethical-social dimension of risk and precaution: *ethical*, in that they concern what counts as harm, who bears responsibility for avoiding it, and what counts as a proportionate response; *social*, in that these judgements are made within a social context, shaped by multiple actors and stakeholders.

That a precautionary measure results from normative judgment is relatively straightforward: choosing how to respond to a danger involves assessing the fitness for purpose of the measure, weighing its costs, and considering alternatives, and is therefore naturally understood as something constructed – resulting from judgement and choices – rather than simply given. Less straightforward is the idea that risk itself is also at least partly constituted by normative judgement, rather than being a pre-given feature of the world.

In the field of risk analysis, risk is often 'objectified' into a purely quantified, factual notion. This fails to capture the value-ladenness of risks – at a minimum, the claim that something is undesirable. But treating a risk as wholly subjective notion, turning it into the outcome of social processes alone, is equally mistaken, as this does not do justice to the fact-ladenness of serious risks. At a minimum, there must be some state of the world that the risk is *about*. Hansson (2010) captures this with his *dual risk thesis*:

> "An accurate and reasonably complete characterization of a risk must refer both to objective facts about the physical world and to (value) statements that do not refer to objective facts about the physical world." (p. 236).

Risks, according to Hansson, thus contain both objective and normative components. The challenge, as he puts it, is to identify the factual and valuational components of risk claims and understand how they combine (p. 237).

If risks are irreducibly combinations of fact and value, they cannot simply be read off the world; someone makes the evaluative judgements that constitute them as risks worthy of concern and response. This raises the question of who makes these evaluative judgements and on what basis. The social constructivist perspective highlights this social aspect of risks without necessarily collapsing into subjectivism. As Burgess (2015) puts it, it directs us "to understand the particular economic, social, political, and cultural influences and actors that lead to the singling out and elevation of some things over others." (p. 4). On this view, a risk emerges through a social process of claims and responses. An analysis of risk thus includes analysing this process, examining both the claims, the claim-makers, and how institutions and people respond (Burgess, 2015: pp. 6-8).

This social constructivist perspective opens the way to a discursive view on risk. Scholars in the rhetoric of risk approach risk as something that comes into being (i.e. is constituted) at the moment of communication (Jensen, 2015, p. 3) and is constantly subject of persuasive endeavours (Danisch, 2010, 2011). Risk thus becomes something that is discursively constructed and, in turn, the product of deliberation (Jensen, 2015).

These perspectives bring risk into view as a fact- and value-laden *idea about the future* that makes a call on the present for precautionary action. In this sense, claiming that there is a risk resembles what the literature on emerging technologies calls a 'vision' (see for example Tepper, 1996; Grunwald 2007). In this literature, visions are commonly understood as ideas about the future of a certain technology and its impact on society. Like risks, visions also include both factual, non-factual (prospective ideas about the future), and evaluative components. And like risks, visions are inherently normative because they outline futures considered worth pursuing (and thus, which futures to avoid). Indeed, as discussed in Chapter 6, the quantum cyber security threat has been identified as one negative component of a broader vision of quantum computing as a geopolitical asset.

Understanding a claimed risk of a future technology as a specific type of vision, as I propose here, has two upshots. First, viewing risk claims as a type of vision brings into view their character as non-neutral ideas about the future that shapes the present. The meanings assigned to a future technology in a vision – what it will be, what it will be able to do, and what it will be able to do good (or bad) for – influence decisions made about its development and regulation. Understanding 'risk' as one specific meaning projected onto a (future) technology, makes clear that calling something a risk is a normative and thus ethically salient act.

Second, and more importantly, the vision perspective comes with an approach for examining the ethical-social dimension of visions. The aim of vision assessment is to examine the meanings attributed to some future possibility and the assumptions embedded therein (Grunwald, 2007; Schneider et al., 2021; Sand, Grunwald, &

Nordmann, 2023; Nordmann & Grunwald, 2023; Mehnert & Grunwald, 2024). This may include epistemic assumptions about the plausibility of some scenario. But beyond investigating the content of visions, their 'authors' must be investigated too. Just as relevant as it is to investigate substantial claims about the future is it to examine *who* is making these claims – and with what possible interests – and *how*, that is, are they openly contestable or at least traceable (see Chapter 6). Risks, seen not as pre-given but as partially socially constructed, have authors too: specific actors frame, certify, and communicate risks, shaping what gets taken seriously and what does not. Adopting a vision perspective thus has a methodological pay-off: it provides an analytical approach suited to examining both the content of risk claims and their authorship.

In Section 4, this understanding of risk as a vision is taken forward to develop a framework for analysing claims to risk and proposals for precaution, drawing on the aspects highlighted here: the normative judgement embedded in any risk claim, the multiple levels of uncertainty that apply to it, and the justification of precautionary action and its proportionality.

The discussions in this section show that the questions raised in the introduction – (i) how to weigh the societal costs of the QSC-transition, (ii) how to calibrate timelines and plans, and (iii) how to distinguish principled from opportunistic appeals to precaution – cannot be answered by technical risk metrics alone. (i) Weighing the different societal aspects of that transition now appears as a matter of deciding which harms and costs are allowed to count within a particular vision of quantum risk, and whose evaluative judgements shape that vision. (ii) Calibrating timelines and plans, in turn, requires more than projecting worst-case scenarios: it must also account for the degree of uncertainty about occurrence, impact, and especially cause, and extend that scrutiny to the risk visions themselves, asking how well-grounded their underlying assumptions are, how explicitly those assumptions are stated, and whose interests shape them. (iii) Principled versus opportunistic appeals to precaution can be distinguished by asking whether the visions that justify precaution make their value-judgements, cost assumptions, and authorship explicit and open to contestation, or whether they instead mobilise the quantum threat in a selective way that obscures who benefits from urgent action.

In Section 5, I return to these three questions in examining the discourse around the quantum cyber security threat, tracing how it articulates societal trade-offs, calibrates timelines, and appeals to precaution in the proposal for a transition to QSC.

4. A framework for assessing risk and precaution as a vision

4.1 Three levels of analysis

Drawing on the preceding discussions of risk, precaution, and their ethical-social dimensions, a framework emerges for the ethical assessment of precautionary action in response to a risk whose inducing cause remains uncertain. Understanding a risk claim[12] and the proposed precautionary response as a specific type of *vision* – a normatively charged idea about a future scenario – means that the method of *vision assessment* offers a useful approach to ethically analyse them. While not a single, established method yet, a three-level approach to the analysis of visions was proposed in Chapter 6: the *substantive level* addresses the content of the claim (i.e. the *meaning* assigned to some future scenario); the *rhetorical level* examines the rhetorical moves made to make the vision normatively compelling; and the *agential level* addresses who is making it, with what aims, and through what process.

In the case of risk and precaution discourse, these levels are analytically interdependent rather than merely parallel. The rhetorical level in particular functions less as a separate object of analysis than as a critical check on the substantive level: it asks whether the way a risk claim is framed accurately reflects what the underlying evidence supports, or whether it instead exaggerates or downplays the risk beyond what that evidence warrants. For instance, describing a speculative future scenario in terms of near-certainty – treating an 'if' as a 'when' – does more rhetorical work than the underlying evidence may justify.

The agential level also interacts with the substantive level: because risk claims are inherently normative and contestable, the process through which they acquire authority is itself ethically significant, and scrutinising that process is inseparable from scrutinising the claims themselves. Distinguishing the three levels is useful because each draws attention to different aspects of a risk discourse: the content of the claim, the argumentative moves that sustain it, and the actors who advance it. Together they provide multiple entry points for ethical scrutiny. Because these levels do not operate independently, however, they serve as a heuristic rather than sequential analytical steps, and in practice often run together.

[12] I use the term 'risk claim' to foreground risk as a statement asserting that a particular technological development will, if it materialises, give rise to a specific harm. When a policy document, for example, states that a cryptographically-relevant quantum computer poses a serious cyber security risk, this statement is doing several things: it asserts that some future technology (a CRQC) will exist, that it will have certain capabilities, that those capabilities will be used in harmful ways, and that this warrants serious attention now. All of this is part of the risk claim. Treating risk as something claimed (rather than simply given) is part of the methodological move to understand risk as a specific type of vision.

4.2 Evaluating questions

As the objective here is not only to describe the vision of risk and precaution, but to ethically assess it, the two main evaluative questions guiding the analysis are:

1) *Can the claimed risk be considered plausible*[13]*?*
2) *Can the proposed precautionary measure be considered appropriate?*

Together, answering these questions serves as a stepping stone to assess whether a specific precautionary response to some claimed risk can be considered ethically defensible. To examine both questions, the three-level analytic scheme of vision assessment is adopted (see Table 1). The substantive level focuses on the content of the claim: What does the risk claim entail, why should the scenario be considered unwanted, and how plausible is its materialisation? For the precautionary response, the substantive level asks what the proposed measure is supposed to do, and what its 'price' is, i.e. the direct, indirect, and opportunity costs.

At the rhetorical level, it is asked how the risk claim and proposed precautionary response are made normatively compelling, prompting action? For the risk claim, this means asking how the risk is framed, and how it is argued that the risk should be taken seriously and acted upon. For the precautionary response, it means asking whether the risk meets the conditions for invoking the precautionary principle, how the price of precaution is justified, and what alternatives exist.

At the agential level, the question is who is claiming the risk or proposing the precautionary measure, and with what aims. For the risk claim, this means asking who is making it, what interests and aims underpin it, and to what extent it is open to contestation and broader deliberation. For the precautionary response, it means asking what interests are at stake in the proposed measure, and to what extent that measure is open to contestation and broader deliberation.

Together, the three levels provide the analytical basis for answering the two evaluative questions: scrutinising the content of the claims, the normative reasoning behind them, and the judgements shaping them helps assess the plausibility of the risk claim and the proportionality of the proposed precautionary response – and, drawing on this, the ethical defensibility of precautionary action.

Table 1 summarises the framework, organising the two evaluative questions as its rows and the three levels of analysis as its columns.

[13] 'Plausible' here does not mean 'likely to happen' but rather 'reasonable to claim some scenario as a risk'.

Table 1. A framework for assessing the ethical justifiability of a risk discourse

Level of analysis *Unit of analysis*	*Substantive level*	*Rhetorical level*	*Agential level*
1) **Can the risk claim be considered plausible?**	What is the claimed risk about?	How is the risk claim made normatively compelling?	Who is claiming the risk and with what aims?
	What does the future scenario of bad outcomes entail? *Why should the scenario be considered undesirable?* *How plausible is the risk across all three types of uncertainty – occurrence, impact, and causal?*	*How is the risk being framed?* *How is it argued that the risk should be taken seriously and acted upon?*	*Who is making the risk claim, and what are their potential interests in doing so?* *To what extent is the risk claim open to contestation and broader deliberation?*
2) **Can the precautionary measure be considered appropriate?**	What does the proposed precautionary response entail?	How is the precautionary measure made normatively compelling?	Who is proposing the precautionary measure and with what aims?
	Does the risk meet the conditions for invoking the precautionary principle? *What is the proposed measure supposed to do?* *What are its direct, indirect, and opportunity costs?*	*How is the price of precaution justified?* *What alternatives exist?*	*Who constitutes the proposed measure?* *What are potential interests in the proposed measure?* *To what extent is the measure open to contestation and broader deliberation?*

In the next section, this framework is put to use to analyse the discourse around the quantum threat to cyber security. Before doing so, it must be noted that the subquestions listed in Table 1 are intended to clarify what each main question concerns rather than to constitute a checklist to be worked through systematically. They

indicate the kinds of considerations that bear on each evaluative question and level of analysis, but the analysis in the next section addresses them in an integrated way, giving more weight to those that are most salient in the quantum case. In other applications of the framework, these subquestions may be complemented with additional considerations as the case requires. The framework thus functions as an orientation for the analysis rather than a strict script for it.

5. Migrating to quantum-safe cryptography: Better secure than sorry?

Mobilising the discussion in Section 3 and the framework developed in Section 4, this section investigates the quantum threat to cyber security as a specific 'risk vision'. The analysis combines description and evaluation: it draws on the technical and empirical literature to characterise the risk claim and the proposed precautionary response as they currently figure in the discourse, and it subjects both to ethical scrutiny using the framework developed above. The aim is to eventually answer the question: is a call for a society-wide transition to QSC ethically justifiable? To investigate this, I first scrutinise the risk claim itself (Section 5.1), and then examine the proposed precautionary response (Section 5.2). In both steps, I look at both the content of the vision, i.e. the substantive level, and the 'visioneering process' itself, including the visioneers, i.e. the rhetorical and the agential levels. In doing so, I follow Table 1.

5.1 The risk: a cryptographically-relevant quantum computer

5.1.1 What is the claimed risk about? (substantive level)

To assess what the claimed risk is about, I first ask: What does the future scenario of bad outcomes entail? (Table 1, substantive level: future scenario). As Section 2 discussed, the quantum threat to cyber security consists in the potential of a large-scale quantum computer to break encryption schemes that are currently widely used to secure digital data. Framing this scenario as a risk amounts to qualifying it as *undesirable*. This assessment appears to be largely uncontroversial. There are many reasons to protect data from unauthorised access, ranging from privacy and intellectual property to protection against fraud and the safeguarding of state secrets. Cryptography is a cornerstone of digital security, and the prospect of widely deployed encryption schemes becoming vulnerable would expose sensitive information and communication to misuse on a large scale. In that sense, the scenario of a CRQC compromising current encryption can reasonably be characterised as an unwanted scenario (Table 1, substantive level: undesirability).

The next question is: How plausible is this risk? How can we justify taking this 'risk' seriously given the uncertainties involved? (Table 1, substantive level: plausibility). Before I discuss the most fundamental uncertainty – whether a sufficiently powerful

quantum computer will ever be built – it is worth first mapping the other levels of uncertainty, since they bear on how serious the risk would be if that condition were met.

Looking at *occurrence uncertainty* – whether the unwanted event would occur given the presence of its cause – the picture is relatively clear. A CRQC is by definition a quantum computer capable of breaking current encryption, so the link between cause and effect is largely definitional. The main residual uncertainty at this level is whether malicious actors would actually exploit this capability, which is an assumption worth making explicit, but not one that substantially undermines the risk claim.

With regard to *impact uncertainty* – uncertainty about the scope and severity of impact – the case also seems to be comparatively strong, although with an important qualification. The type of encryption that is vulnerable to an attack of a CRQC, i.e. public-key cryptography, is deeply embedded in digital infrastructures, securing communication, data storage, and authentication across a wide range of domains (Bernstein et al., 2022, p. 2). More specifically, the cryptographic functionalities that are *de facto* standard for public-key cryptography, such as the RSA cryptosystem and elliptic curve cryptosystems, are precisely the kind of systems that will be no longer secure when attacked by a CRQC (Chen et al., 2016). If these systems were to become vulnerable, the consequences would be significant and potentially very far-reaching – again, presuming that adversaries exploit these vulnerabilities. However, the severity of the impact would depend significantly on the state of digital infrastructures at the moment a CRQC materialises, in particular, how much of the vulnerable infrastructure has by then already been migrated to quantum-safe alternatives. This makes impact uncertainty not just a question about the scope of currently vulnerable systems, but about the state of those systems at the future moment when a CRQC materialises.

Considering *cause uncertainty* – uncertainty about the emergence of the cause of the risk – we enter more unclear terrain. The plausibility of the risk claim ultimately depends on whether it is reasonable to expect that a CRQC will be built ('if'), and if so, on what timescale ('when'). The timeline also matters because if it would be deemed virtually impossible to build it within, for instance, the coming 50 years, this would have implications for the assessment of the risk claim and by extension, to any precautionary response deemed appropriate.

Now, the field of quantum computing is steadily progressing. Recent developments point toward CRQCs becoming more feasible rather than less, with estimates of the timeline getting closer rather than further away (Babbush et al., 2026; Cain et al., 2026). This does not, however, amount to definitive evidence that a CRQC is forthcoming or that its timeline can be reliably estimated (see also Doyle et al., 2026). The number of qubits required to break public-key cryptography remains significantly beyond what current systems have achieved. Moreover, achieving cryptographic

relevance requires more than scaling up the number of qubits: it depends on a range of intertwined engineering and scientific challenges, including breakthroughs in quantum error correction that account for the quality of the qubits. While Shor's algorithm is theoretically proven to solve the relevant mathematical problems, the step from current quantum devices to effective cryptographic capability depends on resolving real challenges that have not yet been overcome. What the progress does establish, at a minimum, is that the prospect of a CRQC cannot be confidently dismissed.

When, if ever, a CRQC will materialise remains one of the most contested questions in the field. One frequently cited source is the Quantum Threat Timeline Report, which in its 2025 edition surveyed 26 experts on the likelihood of a quantum computer capable of breaking RSA within 24 hours. When asked about the likelihood of a CRQC within 10 years, the majority of respondents estimated it at 50% or lower; within 20 years, the vast majority placed the likelihood at 50% or higher; and within 30 years, at 70% or higher (Mosca & Piani, 2026).

These estimates suggest that expert opinion leans toward taking the threat seriously, but they should be interpreted with caution, as the samples are small and the respondents' status as experts in the field may introduce – consciously or unconsciously – a potential bias (see the discussion at the agential level). The literature also reflects no consensus about the feasibility and timelines of large-scale quantum computing (cf. Dyakonov, 2019). Meanwhile, the timeline for quantum computing has been repeatedly revised, prompting the field's own self-critical observation that quantum computers have been 'ten years away' for several decades now (Valsorda, 6 April 2026). This points to something important for question (ii): timelines should not be taken as neutral technical forecasts. They are themselves reflections of visions, shaped by assumptions about the pace and direction of technological progress, and perhaps even by the interests of those who construct them. Treating a timeline as a given rather than as an interpretive claim obscures the normative choices embedded in it and forecloses the kind of recalibration that a more explicit analysis makes possible.

5.1.2 How is the risk made normatively compelling? (rhetorical level)

When it comes to the framing of the risk and of the rationale for action, two rhetorical observations stand out.

The first concerns the framing of causal uncertainty (Table 1, rhetorical level: framing). As the substantive analysis showed, the most fundamental uncertainty surrounding the quantum cyber security threat is causal, i.e. whether a CRQC will ever materialise. Justifying taking the risk seriously thus relies not only on empirical claims, but also on evaluative choices about how to treat uncertainty. The identification of Q-day as a pressing danger is not simply a matter of technical forecasting. From a visions

perspective, framing the emergence of a CRQC not as a *possibility* but as an *inevitability* is not a neutral move. It reflects values such as security, (political) responsibility, and, arguably, a certain technological optimism about quantum computing's trajectory. Treating the emergence of a CRQC not as an 'if' but as a 'when' reduces the salience of the causal uncertainty about the materialisation of a CRQC, and strengthens the normative pull towards action. This directly bears on question (i): how the societal costs of the transition are weighed depends in part on whether the risk is framed as certain or merely possible.

The second concerns the framing of impact (Table 1, rhetorical level: taking the risk seriously). The consequences of a CRQC for cyber security are frequently portrayed as outright disastrous, exemplified by terms like "cryptocalypse" and "Q-day". Such 'catastrophisation' frames the impact as total and immediate, obscuring possible gradations of severity and foreclosing deliberation about proportionate responses. The underlying assumption – that a CRQC will be available to and used effectively by malicious actors – is worth making visible. Doing so allows for a more informed view of the plausibility of the risk, and opens up space for thinking about possible entry points for precaution.

These two rhetorical moves – treating arrival as inevitable and consequences as total – reflect a broader logic of *securitisation*. In short, securitisation refers to the process by which an issue is constituted as primarily a security threat that demands urgent response (Hansen & Nissenbaum, 2009). Hansen and Nissenbaum note that designating something as a security problem endows it with a status and priority that 'non-security problems' do not have (p. 1156). Constituting an issue as a security threat creates a sense of urgency, which compresses space for deliberation – for example, about installing security as the dominant value against which other considerations are subordinated.

The securitisation of the quantum cyber security threat becomes particularly visible when the harvest-now-decrypt-later scenario is invoked as a rationale for immediate action. Without dismissing this scenario as unfounded – its implications for timing are genuine – it is important to recognise how it can operate rhetorically: substituting urgency for argument and bypassing deliberation about timing and alternatives.

The effect of securitisation is further strengthened when the issue is 'technified', i.e. framed as a technical matter requiring expert knowledge rather than a political question open to broader contestation (Hansen & Nissenbaum, 2009, p. 1167). As Barea (2026) observes, this logic of securitisation is visible in the broader discourse on quantum technologies, within which the threat to encryption is presented as one of the main rationales for action.

Identifying these rhetorical moves and the securitisation logic underpinning them does not imply that the risk is necessarily unfounded. Rather, it indicates that any risk

assessment premised on future CRQCs must be transparent about the profound uncertainties surrounding their development, and must explicitly argue why the threat warrants serious precautionary attention despite those uncertainties. Without this transparency, it becomes difficult to distinguish principled from more opportunistic appeals to precaution (question (iii)). Where this justificatory step is omitted – where the possibility of a CRQC is simply assumed rather than argued for – the discourse bypasses an important site of ethical contestation and forecloses questions that deserve to remain open.

5.1.3 Who is making the risk claim and with what aims? (agential level)

Asking who is making the risk claim, I observe that the current discourse is largely driven by experts in quantum computing, cyber security, and related policy domains such as critical infrastructure protection. This is not surprising given the technical character of the subject. As the rhetorical analysis showed, the technical framing of the issue is itself part of how the discourse acquires authority. However, a narrow, technical discourse may foreclose broader scrutiny: by positioning experts as the authoritative voices, it creates conditions in which normative questions – about acceptable levels of uncertainty, or the appropriateness of different precautionary responses – may go unasked. This indicates a potential strategic dimension to the technical framing of the risk (see Table 1, agential level: potential interests).

This matters because technical and normative questions should not be conflated. Technical specialists are well placed to evaluate technical questions about, for example, the feasibility of quantum computing and relevant indicators of significant progress towards that goal. But they are not necessarily in a superior position to address normative questions that are equally at stake: about how much uncertainty is acceptable in treating a scenario as a risk, and whether it warrants precautionary action. Answering these questions involves normative judgement about acceptable levels of risk, appropriate responses to uncertainty, and the allocation of resources. As such, they should be made explicit and open to broader scrutiny rather than left to resolve by default within a technical discourse. However, given the predominantly technical character of the discourse, the risk claim is currently subject to scrutiny mainly within expert communities (see Table 1, agential level: risk claim open to contestation and broader deliberation).

This dynamic cannot be seen separately from the aims of those advancing the quantum cyber security risk claim. The most directly stated aim is of course harm prevention: those raising the alarm are, in the first instance, seeking to avert what they genuinely regard as a serious threat. Yet actors involved in the development and study of quantum technologies may also have other institutional or professional incentives to emphasise the significance of the threat, for instance to secure funding or promote particular research agendas. At the same time, these incentives are counterbalanced by

reputational risks associated with overstatement. The resulting picture is not one of bias in a simple sense, but of a field in which claims about the future are shaped by a combination of expertise, interests, and institutional contexts.

This discussion also bears on question (iii): if normative questions are resolved by default within a narrow expert discourse, the appeal to precaution becomes harder to distinguish from one driven primarily by particular interests rather than by transparent ethical reasoning.

5.2 The precautionary response: implementing quantum-safe cryptography

5.2.1 What does the precautionary measure entail? (substantive level)

In response to the cyber security risk associated with future CRQCs, policymakers and experts have called for a transition to cryptographic standards that are resistant to attacks by quantum computers, so called quantum-safe cryptography (QSC) (Table 1, substantive level: proposed measure). This response is clearly precautionary in nature: despite uncertainty about whether and when a CRQC will materialise, it seeks to mitigate potential harm in advance. In terms of the earlier distinction, the discourse reflects a preference for a false positive (acting unnecessarily if the risk does not materialise) over a false negative (failing to act when the risk materialises).

A first step in assessing the appropriateness of this response is to examine whether the case meets the conditions for invoking the precautionary principle (Table 1, substantive level: invoking the precautionary principle). For this, Gardiner's framework is useful. The first condition – significant uncertainty about probabilities or outcomes – appears satisfied. While the consequences of a CRQC are relatively well understood, there is substantial uncertainty about the occurrence of the risks associated with this, due to uncertainty about the feasibility and timeline of a CRQC.

Gardiner's second condition – low concern for potential gains and high concern for protecting a minimum – requires more careful interpretation. Since quantum computing promises significant benefits, there is a concern for gains. Precautionary measures that would halt or severely restrict its development, such as a moratorium, would thus come with considerable opportunity costs (Table 1, substantive level: the costs of precaution). However, the transition to QSC does not foreclose these gains, as this specific precautionary measure does not demand refraining from (or slowing down) the development of a large-scale quantum computer.[14] Rather, it aims to preserve a

[14] In emphasising that the specific precautionary measure under consideration does not foreclose the gains associated with the risky activity, I am interpreting rather than strictly applying Gardiner's second condition. Gardiner's condition concerns the decision-maker's general orientation toward the gains of the risky activity itself. My move shifts the focus to whether those gains are actually foregone by the

critical minimum – namely, the integrity of digital security – while allowing technological development to continue. Given the widespread use and significant role of cryptography in modern society, concern for protecting this minimum can plausibly be considered high.

The third condition – that alternative options would have outcomes that are difficult to accept – also appears to hold, albeit with an important qualification. If the risk of a CRQC is taken to be non-negligible, then maintaining the status quo would imply accepting the possibility of large-scale compromise of digital security. At the same time, this judgement depends on how the underlying uncertainty is assessed. The case for precaution is therefore not independent of the earlier evaluation of the risk claim. This illustrates why, as question (ii) suggests, timelines and plans need to be continuously recalibrated rather than fixed at a particular moment in time.

Taken together, Gardiner's conditions appear to be satisfied for some form of precautionary action. However, which action is warranted depends heavily on how one weighs the societal costs and benefits involved, which is raised in question (i). Establishing that precaution is warranted in principle does not yet establish that the specific response – a transition to QSC – is appropriate. The next question is therefore one of proportionality.

Here, Munthe's idea of the price of precaution is useful. The QSC transition involves significant direct costs in the form of resources needed to update or replace cryptographic systems (see Section 2). The most apparent indirect costs lie in the introduction of new vulnerabilities associated with relatively immature post-quantum standards. Finally, there are opportunity costs: resources devoted to QSC are not available for addressing more immediate cyber security threats or other societal priorities.

At first glance, these costs may seem easily outweighed by the potential consequences of a "cryptocalypse". Yet the trade-off is more complicated than it first appears. The most fundamental complication is the causal uncertainty discussed above: if there is genuine uncertainty about whether a CRQC will ever materialise, then the proportionality of an immediate and costly transition cannot be assessed simply by comparing its price against the severity of the worst-case outcome. The probability – or more precisely, the argued plausibility – of that outcome matters too. A transition that appears obviously justified under the assumption of near-term inevitability looks considerably less clear-cut if the prospect of a CRQC remains genuinely uncertain.

Moreover, timelines also matter, though in a distinct way: a longer horizon would allow for a more gradual and adaptive response, potentially at lower cost, whereas a

chosen precautionary measure. I take this to be a defensible adaptation given the specific structure of this case.

compressed timeline leaves less room for recalibration. In both respects, the proportionality of the price of precaution depends on normative judgements about what level of uncertainty justifies what level of precautionary investment. This shows the interrelatedness of questions (i) and (ii): the weighing of societal costs of the QSC-transition must run together with the fine-tuning of timelines and options for action.

At the same time, certain considerations strengthen the case for early action. The 'harvest-now, decrypt-later' scenario implies that the effective window for precaution is significantly shorter than the timeline for a CRQC itself. Under this scenario, the risk is not a future one anymore but already applies to the present, as data collected today may be decrypted in the future. This compresses the timeline for action and strengthens the case for earlier rather than later precaution, even under conditions of significant causal uncertainty. This responds to question (ii), in the sense that it provides grounding for timelines for action.

Finally, the assessment of proportionality must take into account alternative responses, which is central to question (i) about the weighing of societal costs. The relevant comparison is not simply between action and inaction, but between different forms of precaution. The current discourse tends to frame the transition to QSC as *the* precautionary response, as though it were the only option on the table. But proportionality can only be evaluated comparatively, relative to both the risk and the available alternatives. So, what are the alternatives?

A moratorium on quantum computing is not straightforwardly realistic: the technology is being developed across multiple countries and a moratorium would require international coordination that is not necessarily politically plausible or even realistic, especially given the geopolitical framing of quantum computing as a strategic asset and the fact that, as with other dual-use technologies, a moratorium accepted by all countries appears to be unlikely. Moreover, it would foreclose genuinely valuable research and potential benefits that can be reaped from it. While this would not necessarily rule out a moratorium as a defensible option, it would require a strong justification given the extent to which it forecloses valuable gains.

Another option could be a staged response, combining active monitoring of quantum computing developments with pre-set technical milestones that trigger scaling up precautionary measures. This would allow the intensity of the response to track the evolving plausibility of the threat rather than committing to a full transition now under conditions of profound uncertainty. This option also has a real weakness: the harvest-now-decrypt-later scenario means that waiting for clearer evidence before acting may already be too late for data being collected now. The asymmetry between the long transition timeline and the threat to current data makes a 'wait and see' approach genuinely risky.

While these may not be the only conceivable alternatives, QSC does have something going for it that they lack: it addresses both the future threat and the present one through the harvest-now-decrypt-later scenario, and it does not require restricting a technology whose international development is difficult to coordinate. This goes some way toward explaining why QSC has emerged as the dominant response. However, even within the QSC transition itself there is a spectrum of approaches – from full immediate migration to phased migration to crypto-agility strategies that build in flexibility for future cryptographic updates – whose relative merits deserve open deliberation. A precautionary response that is adopted without this kind of comparative scrutiny cannot be fully justified as proportionate, regardless of whether it turns out to be the right choice.

A precautionary response adopted without comparative scrutiny cannot be fully justified as proportionate, regardless of whether it turns out to be the right choice. This directly bears on question (iii). Where alternatives go unexamined and the choice of response is left implicit, it becomes difficult to distinguish a principled appeal to precaution from one shaped primarily by institutional momentum or the default assumptions of those proposing it.

5.2.2 How is the precautionary measure made normatively compelling? (rhetorical level)

Three rhetorical observations are worth making about how the case for the QSC transition is constructed and justified (Table 1, rhetorical level: justification). First, the harvest-now-decrypt-later scenario does important argumentative work: by establishing that adversaries may already be collecting encrypted data in anticipation of future decryption, it brings the future threat into the present. This makes delay appear immediately costly rather than merely cautious. This may be a legitimate move, assuming that the scenario is real, but its effect is to compress the space for deliberation about timing and alternatives, which is worth highlighting.

Second, certain steps of the transition have been framed as 'no-regret' measures. This applies, for example, to maintaining a cryptographic asset inventory (European Commission, NIS Cooperation Group – PQC Workstream, 2025, p. 11). Framing these interventions as no-regret presents them as beneficial regardless of whether the threat materialises. Applied selectively to genuinely context-independent measures, this framing is defensible, though the urgency might be contestable. Moreover, it should not obscure that the further steps toward the transition carry a more demanding justificatory burden.

Third, the framing of QSC as the obvious precautionary response is itself a specific framing. that, while plausible, risks foreclosing serious comparative evaluation (Table 1, rhetorical level: alternatives). Presenting something as obvious performs

argumentative work: it substitutes the appearance of consensus for explicit justification. It makes comparative evaluation appear unnecessary, instead of concluding that QSC is the best strategy after such evaluation. This reflects what Berling (2011) calls 'objectification', when scientific consensus is mobilised to suppress debate about what are essentially political and normative choices. The point is not that better alternatives exist and are being ignored, but rather that genuine justification of proportionality requires comparative evaluation against alternatives, even if QSC is likely to ultimately emerge as the best strategy. This is directly relevant to questions (i) and (ii): if QSC is treated as the self-evident response, the weighing of its societal costs and the consideration of alternative pathways are omitted by assumption rather than by argument.

5.2.3 Who is proposing the precautionary measure and with what aims? (agential level)

The need for genuine comparative evaluation is strengthened when considering that the discourse about the quantum cyber security threat and the implementation of QSC as the chosen precautionary path is driven by a relatively limited set of actors (Table 1, agential level: actors): standardisation and regulatory bodies like NIST, ENISA and the European Commission, national governments and security agencies, and an expanding industry of quantum-safe solution providers.

Each operates within institutional contexts that shape their perspectives and priorities (Table 1, agential level: potential interests). Industry actors have commercial interests in the development and adoption of post-quantum solutions. National security agencies operate from a mandate to take threats seriously, which can create a structural tendency toward urgency and precaution. Standardisation bodies like NIST, while committed to technical rigour, work in close collaboration with national security communities, which shapes which problems are prioritised and how solutions are framed.

This does not invalidate the proposed response, but it does mean that the framing of QSC as the obvious and urgent precautionary path deserves scrutiny that goes beyond technical validation. The choice of which precautionary measure to adopt, and when, is a normative decision that should not be determined by a particular set of (security-driven) institutional priorities that may not fully represent broader societal considerations (Table 1, agential level: contestation and deliberation). This is not unlike the position of virologists advising during the Covid-19 pandemic: while indispensable for understanding the risk and possible responses, they were not positioned to have the final say on how that risk should be traded off against other societal interests and values. The same distinction applies here: however technically competent the experts may be, questions that are deeply normative should not be resolved are pure technical issues.

While the complexity and stakes involved may well justify action coordinated 'from above', the question is how the normative trade-offs are resolved. Decisions about how to weigh the costs of the QSC transition against competing societal priorities, what level of uncertainty justifies what level of precautionary investment, and which values should be prioritised are not pure technical questions but deeply normative ones. Where public resources and governmental authority are involved, for example through mandating a transition to QSC, these trade-offs should be openly justified.

5.3 Plausible risk, appropriate precaution?

So where does this leave us? What this analysis establishes is, first, that assessing the plausibility of the claim that a future CRQC poses a cyber security risk is complicated primarily because of the profound uncertainty about its materialisation. Second, it reveals that such cause uncertainty complicates the assessment of the QSC's proportionality of the QSC transition, shifting the burden from evidence to explicitly normative argumentation: where evidence cannot settle the question, the reasoning must do more work. In the current discourse this difficulty tends to remain implicit rather than explicitly acknowledged, with the 'if' often quietly becoming 'when'. This is an ethically salient omission because it bypasses a necessary step in the justification of both the risk claim and the precautionary response it motivates.

The analysis in this chapter forms a start of addressing this gap. It points to the conclusion that, given that confident dismissal of a CRQC is increasingly difficult to justify and that the consequences of a large-scale cryptographic failure are potentially serious, widespread, and difficult to reverse, the QSC transition is ethically defensible as a proportionate precautionary response. However, making its justification explicit reveals that this proportionality is conditional: it depends on how one weighs profound uncertainty about a CRQC's plausibility against the very real and immediate costs of the transition. Revealing the justificatory steps underpinning the current discourse thus serves a dual purpose: it provides that narrative with more explicit ethical grounding, while simultaneously opening it to scrutiny by making the underlying normative reasoning visible.

The analysis also foregrounds that this reasoning comes from somewhere. Visions have authors, and understanding risk as a negative vision means that it matters who claims that there is a risk and who proposes a precautionary response for evaluating its ethical defensibility. The current discourse is dominated by a limited set of actors, mostly regulatory bodies, security agencies, standardisation bodies, and industry players. The normative assumptions embedded in the risk claim and the precautionary response – about what level of uncertainty suffices to act and how to weigh the price of precaution against the perceived price of inaction – are therefore being resolved from a particular vantage point. Moreover, within the resulting predominantly technical and security-

centred discourse, they are not always recognised as normative questions at all. Highlighting this agential dimension does not invalidate the dominant narrative, but it strengthens the case for scrutinising it.

6. Conclusion

Large-scale quantum computers would provide a tool to break widely used encryption methods, posing an unprecedented cyber security threat to our digitalised societies. Such machines do not yet exist, and their feasibility is not uncontested. Nevertheless, a discourse has taken shape that strongly calls for a serious precautionary response to this potential threat, in the form of migrating to encryption methods resistant to quantum attacks. Such a transition demands investing real resources in anticipation of a speculative risk. Recognising that precautionary action always demands justification – and that the costs and uncertainties involved in this case make that demand particularly pressing – this chapter asked: When is it ethically defensible to treat an uncertain technological development as a risk that warrants precautionary action?

By investigating this, the chapter took up a question that Chapter 5 left open. There, it was argued that for early-stage technologies such as quantum computing – characterised by profound uncertainty about their eventual form and impact – ethical engagement with possible consequences is generally premature: outcomes-oriented scrutiny presupposes a degree of concreteness that immature technologies typically lack. The quantum cyber security case was identified as challenging this default, since the potentially severe consequences of a future CRQC may provide grounds for outcomes-oriented engagement even at this early stage. This chapter developed the conditions under which early-stage engagement with the outcomes of a still-speculative technology may be defensible.

I defined a course of action as ethically defensible when its justification is ethically robust, i.e. when the underlying reasoning is explicit and capable of withstanding scrutiny and contestation across relevant ethical dimensions. In the case of precautionary action in response to a future risk, I argued that these dimensions are the plausibility of the risk and the proportionality of the precautionary measure. More specifically, I argued that the ethical defensibility of treating a future technology as a risk and acting on it by precaution requires:

- that the risk claim is made sufficiently *plausible* through explicit argumentation rather than tacit assumption
- that the precautionary response is shown to be *proportionate* through a transparent weighing of costs, alternatives, and uncertainties

The ethical robustness of this justification, moreover, requires that the normative judgements embedded in both the risk claim and the precautionary response are made

visible and open to contestation. This requirement is especially pressing when the technology at hand is still speculative: the thinner the evidential basis, the more weight is carried by normative assumptions, and the more important it becomes that those assumptions can be examined and challenged.

Drawing on these requirements, I developed a framework for assessing two things: the plausibility of the claim that a future technology poses a risk worth taking seriously, and the appropriateness of the proposed precautionary response. In the case of still-speculative technologies, the uncertainties surrounding both risk claims and precautionary proposals shift the argumentative burden onto assumptions and normative judgements. To account for this, I proposed to approach both as a special instance of a technological vision, i.e. an idea about the future of a technology that can motivate present action. Drawing on the method of vision assessment (Chapter 6), it was argued that a specific risk discourse is best analysed across three levels: investigating the risk claim and the precautionary response itself, the primary argumentative moves to make them normatively compelling, and the specific vantage points of the discursively dominant actors.

The analysis largely supports the narrative that future quantum computers constitute a risk worth taking seriously despite profound uncertainties, and that a transition to quantum-safe cryptography is a plausible and proportionate precautionary response. However, it also revealed that this assessment rests on assumptions that require explicit justification. The first is that a cryptographically relevant quantum computer is sufficiently plausible to warrant precautionary action despite what this chapter calls 'cause uncertainty': uncertainty not about the risk itself but about whether its inducing cause will ever materialise. The second is that alternatives to the QSC transition are implausible or insufficiently effective as proportionate precautionary responses.

The justifications for these assumptions developed here should be taken as a first attempt rather than a final verdict. That the current discourse is largely driven by a limited set of technical and security-oriented actors – in which normative questions are sometimes treated as technical ones – underscores the need for broader scrutiny of those justifications: from those positioned to assess the technical assumptions on their own terms, and from policymakers and other societal actors to weigh the normative trade-offs involved.

The goal of the chapter was never to argue that the dominant narrative is wrong. The aim was to hold the reasoning that underpins the risk claim and the proposed precautionary measure to a higher justificatory standard than it currently meets, in order – if indeed successful – to render it ethically defensible. By making the implicit justificatory steps explicit, the analysis not only provided ethical grounding where it supports the dominant narrative, but also outlined the openings for contestation by revealing the underlying assumptions and normative trade-offs that remain implicit in

the current discourse. Being more open about the uncertainties and assumptions involved, ultimately makes the discourse about quantum's cyber security threat epistemically more honest and reflexive, more open to ethical contestation, and arguably more resilient to the unexpected.

More concretely, this analysis sought to illuminate three interrelated research questions. With respect to question (i), how to weigh the societal costs of the transition, the analysis shows that the QSC transition is not a purely technical matter but depends on normative judgements about which harms and costs count, and that it matters whose evaluative perspectives shape that assessment. The analysis also identified specific framings in the discourse – such as the shift from 'if' to 'when', the catastrophisation of impact, and the securitisation of the risk – that do argumentative work. Making these normative dimensions and the framings through which they operate explicit is a precondition for weighing the societal costs carefully, and provides entry points for further scrutiny of the assumptions on which the dominant narrative rests.

Regarding question (ii), how to calibrate timelines and plans, the analysis cautions against treating timelines as neutral technical forecasts. Quantum computing timelines have been repeatedly revised. They reflect the state of the science as much as assumptions and aspiration about technological progress and should be read accordingly. Acknowledging this creates room for plans to be adjusted as the situation evolves, instead of being locked in by a particular projection.

With respect to question (iii), distinguishing principled from opportunistic appeals to precaution, the analysis draws attention to the interests of both scientists and industry actors to stress the plausibility of the risk and of the transition to QSC as the solution, alongside the legitimate concerns about cyber security. Where the urgency of the transition is amplified by actors who stand to benefit from it, and where QSC is framed as the self-evident response without comparative evaluation, the boundary between principled precaution and opportunistic framing becomes difficult to discern. By analysing the assumptions and interests embedded in the precautionary discourse, the analysis suggests that the case for precaution is largely principled; meaning, defensible on the ground of explicit ethical reasoning. At the same time, it highlights that this reasoning rests on assumptions and normative trade-offs that are open to contestation.

These questions, however, remain open to deliberation, and require continuous revision in light of new technical developments. This chapter has sought to inform that deliberation.

Chapter 9

Conclusion and discussion

> "Whooping with delight, we slide ever faster down the slide called progress, without asking ourselves in what direction we are going."
>
> Wijsheid, Paul Verhaeghe (2025, p. 51)

1. The epistemic conditionality of ethics

This concluding chapter synthesises the contributions of the individual chapters and traces these insights back to the research questions this thesis set out with. This should provide a clearer view of epistemic conditions that enable a meaningful ethical discussion of a new or emerging technology such as quantum. Or, in other words, how knowledge, and the limits thereof, shape the ethics of technology. The chapter closes with a brief reflection on the research and identifies questions set aside here that point toward avenues for further research.

2. Research questions

This thesis was prompted by the emergence of quantum technologies, a new technological family whose ethical study raises questions that apply more broadly to all new and emerging technologies. While the ethics of technology typically focuses on the ethical questions a new technology raises, this thesis took a step back and asked what enables a meaningful ethical discussion in the first place. Motivated by the specific epistemic context of new and emerging technologies – shaped by limited knowledge and profound uncertainty – it took an epistemic approach to this question. It put central the following research question:

> *What are the epistemic conditions of meaningful ethical engagement with technology?*

In this thesis, two such conditions were formulated: sufficient understanding and appropriate timing. Sections 3 and 4 elaborate on each of these conditions respectively.

Quantum technologies exemplify what the literature refers to as *new and emerging technologies*. The ethics of new and emerging technologies is defined by its anticipatory ambition: by engaging with a technology while it is still taking shape creates an opportunity to influence its trajectory before it becomes entrenched. Engaging with a

technology 'upstream' is an effective way of influencing its consequences 'downstream'. Yet as the Collingridge dilemma articulates (Collingridge, 1980), the price of this opportunity is a challenging epistemic situation, characterised by novelty and uncertainty. I presented these aspects as the twofold epistemic challenge of upstream ethics.

The *challenge of novelty* consists in the need to come to grips with a technology that is new and unfamiliar. But what does 'coming to grips' involve? What does it mean to have (sufficient) understanding of a technology?

In the absence of a comprehensive philosophical account of what understanding amounts to with regard to technology, the first part of this thesis asked:

> *What does understanding mean in relation to technology, and which form of understanding best enables meaningful ethical analysis?*

The thesis showed that any (successful) engagement with technology, including ethical engagement, requires a form of understanding. It was argued that one specific type of understanding – grasping a technology's potential functions – is especially critical for asking meaningful ethical questions. That is, questions grounded in what the technology is and may do, rather than being carried away by assumption-driven speculation. Section 3 summarises this account, as developed in Chapters 2–4, and identifies *sufficient understanding* as the first condition of meaningful ethical engagement with technology.

The requirement of sufficient understanding is especially challenging in the context of new technologies, where such understanding has yet to be developed and is necessarily limited by the early stage of development. This introduces to the second epistemic challenge central to this thesis: *the challenge of uncertainty*. Novel technologies are surrounded by uncertainties of all kinds, about their feasibility and timelines, and about the purposes for which they will be used and the effects thereof. In other words, at the earliest stages of development, technological understanding is inherently underdeveloped. This confronts ethical engagement with a difficult balance: waiting for impacts to reveal themselves risks being too late to intervene effectively, while speculative overreach risks being too soon, drawing attention and resources to scenarios that may never materialise. While several approaches have been developed to navigate this balance, no account existed to assess their relative strengths and weaknesses, and how these shift as a technology develops. The second part of this thesis therefore asked:

> *How should we calibrate our ethical engagement to what is – and is not yet – understood about a technology?*

It was argued that not all forms of ethical scrutiny are equally suited to all stages of technological development. The timing of ethical engagement informs the type of

ethical scrutiny that can be meaningfully pursued. The ethical approach should thus be aligned with the level of technological maturity. This idea of 'ethics readiness' and its implications for ethical engagement with (quantum) technologies are summarised in Section 4.

The investigation into these two epistemic challenges – novelty and uncertainty – and the responses developed in this thesis – sufficient understanding of technology and appropriate timing of ethics – also revealed their entanglement. Articulating what it means to understand a technology illuminates the core of the uncertainty problem: the challenge of conducting ethics under conditions of uncertainty is, at its heart, a challenge of limited understanding. And vice versa: a lack of technological understanding is an ethical problem. When a technology is at an early stage of development, the understanding needed for meaningful ethical engagement is not yet fully available, which is precisely what makes the epistemic situation uncertain. The two conditions are therefore not independent: the account of understanding developed in Part I directly informs the uncertainty problem taken up in Part II.

3. Sufficient understanding of technology

The first part of this thesis has developed a philosophical account of what understanding entails in relation to technology – a topic that had received little to no systematic attention in the philosophy of technology. The account built on the philosophy of science, where understanding has been extensively theorised, and developed an analogue for the technological domain (Chapter 2). The basic idea is that understanding, in both science and technology, is a cognitive skill: the ability to use a tool – a theory or a technology – to achieve an aim. It entails a genuine grasp of the tool, recognising how it can be used successfully. Technological understanding, then, is the cognitive skill involved in the use of technology: it is *the ability to achieve an aim by using a technological artefact*. 'Use' is here deliberately construed broadly to cover a range of activities: designing and operating the artefact, as well as envisioning new applications.

An important feature of this account of technological understanding is that it comes in both degrees and different kinds. The account develops technological understanding as a fundamentally *contextual* concept (Chapter 3). The relevant form and degree of understanding depend on the context of engagement: designing a technology, operating it, or envisioning new applications each requires a distinct type of understanding, focusing different aspects of the technology. The designer needs to grasp the technology's inner workings; the operator its direct consequences; the innovator its functional capabilities. Rather than treating degrees of understanding as varying along a single scale on which some people simply score higher or lower than others, this pluralistic and contextual view recognises distinct forms of understanding suited to

distinct practices. It thereby provides a more precise basis for determining what kind of understanding is required in a given situation.

In response to this thesis's question about understanding and ethics, it was argued that understanding a technology's functional capabilities – what it can do and what it may be used for — is most directly relevant to ethical engagement (Chapter 4). Such functional understanding was initially linked to the context of innovation, where it enables devising new applications, but was also argued to be directly relevant to ethical discussions about technology. Ethics of technology is ultimately concerned with what a technology does in the world: how it may be used, by whom, for what purposes, and with what consequences. Answering these questions requires grasping a technology's potential uses and applications, which is exactly what functional technological understanding provides.

Does that mean that the other forms of technological understanding have no role in ethics? Certainly not. Both the more technical type of understanding, which is primarily relevant to the context of design, and the operational type, which is involved in direct use, can bring ethically salient aspects into view and inform ethical enquiry. The functional type of understanding is even to some degree grounded in these other types of understanding. But this grounding need not reside in the same person; functional understanding can legitimately build on expertise held by others. Moreover, 'technical understanding' alone is insufficient for ethical analysis: knowing how a quantum computer processes information does not by itself tell you how it might be deployed, who stands to benefit or be harmed, or what societal implications follow. Functional understanding bridges this gap by providing a grasp of what the technology affords – what it can be used for – which is the more direct basis for tracing potential impact and identifying ethical stakes.

Furthermore, while functional understanding was argued here to be the most directly relevant type for ethical engagement, it is not claimed to be sufficient on its own. Meaningful ethical discussion also requires an understanding of the context in which a technology is or may be used, familiarity with ethical concepts and methods, and arguably a form of ethical expertise. These questions about what else is needed – and about who we expect to have it – are taken up in Section 5.1.

Taken together, these chapters make three connected contributions. Theoretically, they provide the philosophy of technology with a systematic account of technological understanding. The account highlights the cognitive skill involved in engagement with technology and distinguishes understanding from knowledge or technical skill, situating it within a broader account of human-technology engagement. Practically, the account provides orientation for efforts to promote understanding of technologies like quantum amongst those concerned with its ethical and societal implications: rather than focusing on the underlying physics, such efforts should prioritise getting functional capabilities

– what a technology affords – into view. And finally, they contribute to the ethics of technology by making explicit what has often remained implicit: that the quality of ethical discussion depends, in part, on the quality of understanding that informs it. Sufficient understanding of a technology's functional capabilities underpins meaningful ethical engagement, and articulating what such understanding entails is itself a step towards a more epistemically grounded ethics of technology.

If technological understanding is essential for meaningful ethical engagement – as the first part of this thesis concluded – then limited understanding at the earliest stages of development translates directly into a challenge for ethics. This challenge was central to the second part of this thesis.

4. Appropriate timing of ethics

How can we engage with the ethical aspects of a technology still shrouded in profound uncertainty, avoiding both the speculative overreach of 'hallucinating ethics' and the risk of coming in too late, leaving ethics as an afterthought? How is meaningful ethical engagement possible when technological understanding is necessarily limited? The second part of this thesis explored these questions as presenting the central challenge of 'upstream ethics': ethical engagement at a very early stage of technological development. It developed a theoretical framework for navigating this challenge and put it to work in the ethical study of quantum technologies.

Central to this framework is the concept of ethics readiness (Chapter 5). Ethics readiness refers to a technology's suitability – how 'ready' it is – for specific types of ethical scrutiny. It captures the idea that the suitability of different types of ethical engagement depends on a technology's stage of development. As a technology matures and takes shape more concretely, it becomes ready for increasingly specific ethical questions. Conversely, when a technology is still in a conceptual or early experimental phase, technological understanding is still rudimentary, which limits the questions that can be meaningfully asked about it.

This dynamic has direct implications for the appropriateness of different ethical approaches. Meaning-oriented approaches – focused on the meanings, visions, and expectations attributed to a technology – are less dependent on a technology's concrete shape and are therefore better suited to deal with the profound uncertainties of early-stage development. Outcomes-oriented approaches – focused on the implications of a technology's design and deployment – require more detailed knowledge and gain both feasibility and relevance as a technology matures. The resulting framework informs which ethical questions are best asked when.

Applied to quantum technologies, the ethics readiness framework points towards differentiated engagement. Quantum sensing, being a more mature technology of which there is considerable understanding, allows for outcomes-oriented questions. Quantum

computing and communication, on the other hand, are still at early experimental stages and our understanding of them remains limited. This calls for a meaning-oriented approach that focuses on the visions and expectations currently shaping their development rather than on speculative impacts that cannot yet be concretely assessed.

Responding to this call, the thesis adopted a meaning-oriented approach to the ethical study of quantum computing (Chapter 6). By investigating visions of quantum computing, the focus was shifted from analysing a future technology to studying ideas in the present. Through an empirical analysis of vision statements by actors playing a prominent role in the quantum computing ecosystem, four main visions were identified: quantum computing as a contributor to societal benefit, as a scientific tool, as an economic asset, and as a geopolitical advantage. Each of these visions frames quantum computing as a powerful tool that will bring about change, but differs in the domain in which it locates that change.

After scrutinising these visions at the level of their content, rhetoric, and authorship, several issues surfaced that warrant attention: a pervasive techno-optimism paired with technosolutionist reasoning; promissory language used to legitimise innovation agendas; and a discourse shaped by the interests of a relatively narrow group of powerful actors. Together, these findings yield at least two observations.

The first is that the speculative character of the discourse reflects the still limited understanding of quantum computing. When functional capabilities are not yet well understood, it becomes easier to project optimistic assumptions onto the technology and harder to assess critically whether proposed applications are genuinely feasible or fit for purpose. Incomplete functional understanding, in other words, provides fertile ground for techno-optimism and technosolutionist reasoning. Advancing such understanding would be one way to address these tendencies. In the meantime, as long as that understanding remains limited, quantum visions should be read cautiously as projections shaped by assumptions rather than as reliable forecasts.

The second observation concerns the vanguard and elitist character of the current quantum discourse: the direction and priorities of quantum development are currently being set by a small elite – predominantly large technology companies and a handful of powerful states – with limited input from broader society. This concentration is partly a reflection of the fact that technological understanding of quantum technologies remains largely confined to expert communities. The elitist character of the discourse raises questions about whose values and interests are being inscribed into these technologies, and underscores the importance of a wider societal debate about the desirability and direction of quantum development.

Against the backdrop of this speculative character of the current quantum discourse and its elitist flavour, the notion of Quantum for Good (Q4G) invites scrutiny (Chapter 7): Good for what and defined by whom? As an ambition to mobilise quantum

technologies for societal benefit, Q4G reflects a technosolutionist logic: starting from technological potential and seeking societal problems to match it. Technosolutionism is not inherently problematic, but it does carry ethical risks. To examine whether and in what respect a specific technosolutionist approach is ethically problematic, a set of guiding questions was proposed. These questions probe the construction of the problem, the sociopolitical motivations and criteria underlying the proposed solution, its temporal scope and broader impacts, and the plausibility of the technological approach.

While Q4G represents a welcome shift towards more socially engaged innovation, Q4G's technology-first orientation carries a real technosolutionist risk: social challenges may be reframed in narrow, technically tractable terms, depoliticising inherently political questions and marginalising more effective alternatives. Moreover, activities under the banner of Q4G often rest on speculative assumptions about quantum's future capabilities rather than on a grounded understanding of what these technologies will ultimately be capable of. This risks making the 'for good' label more rhetorical – potentially even misleading – than a genuine commitment. For Q4G to fulfil its promise, and become more than a moral veneer for innovation agendas, it must both be honest about the uncertainties surrounding quantum development and engage with the limits of technological fixes. That means recognising the social and political dimensions of the challenges it seeks to address, and actively inviting debate about what quantum technologies can realistically contribute.

The thesis also examined what might be called the 'quantum for bad' scenario: the potential of a full-scale quantum computer to break widely used encryption methods, posing an unprecedented cyber security threat (Chapter 8). This scenario embodies the most tangible and potentially urgent ethical implication of quantum computing. While no such machine currently exists and its feasibility is not uncontested, urgent calls for precautionary action are being made. This sits in tension with the ethics readiness framework, which suggests that engaging with outcomes – such as a concrete risk – is premature at this stage of quantum computing's development (Chapter 5). When understanding of a technology is still provisional, engaging with its projected outcomes risks treating scenarios as serious threats while these may never materialise.

The case of quantum's threat to cyber security raises the question under which conditions, given limited technological understanding, engaging in outcomes-oriented ethics at an early stage of technology development can be justifiable. More specifically, it raises the question of when it is ethically defensible to take a risk posed by a still-speculative technology seriously as warranting precautionary action.

It was argued that, at early stages of technological development, outcomes-oriented ethics and precautionary action are defensible under specific conditions: the claimed risk must be made sufficiently plausible, and the precautionary response must be

effective and proportionate. This can be read as minimum baseline for technological understanding: a sufficient grasp of what the technology is capable of to evaluate whether the risk claim is credible and the proposed response appropriate. A framework was developed for assessing both and applied to the quantum cyber security case. The analysis largely supported the dominant narrative: the quantum cyber security threat is serious enough to warrant precautionary action, and the transition to quantum-safe cryptography is a plausible and generally proportionate response. By making the implicit reasoning explicit, the analysis provided the ethical grounding the discourse currently lacks. At the same time, it revealed that the plausibility of the risk – and by extension the proportionality of the response – rest on assumptions that often remain implicit. Making these explicit opens them up to the broader scrutiny they deserve.

Taken together, the chapters in the second part of this thesis deliver two inter-related methodological contributions. The ethics readiness framework provides practitioners in the ethics of new and emerging technology with a meta-framework for making informed methodological choices, assessing which approach to ethics is most appropriate given a technology's epistemic context. The method of ethical vision assessment provides a concrete empirical method for meaning-oriented ethics, applicable beyond quantum to any emerging technology. Together, these tools help asking the right questions at the right moment. Finally, through its analyses of visions, ambitions, and fears, this part directly contributed to the nascent ethical discussion of quantum technologies.

5. Reflections and further research

Over the past four years, I have watched the ethical discussion of quantum technologies develop, from an early responsible innovation discourse to a more focused engagement with ethical implications, on both a societal and geopolitical scale. I have joined many workshops, conferences, seminars, and roundtables, and let these discussions shape my research questions. One issue, however, kept surfacing as an undercurrent in these sessions without finding a place in this research project: the role of the ethicist. The question of who the ethicist is and what is expected of them connects to this thesis's topics of technological understanding and early-stage ethics, and points toward two avenues for further research.

5.1 Division of epistemic labour

The first avenue concerns the *division of epistemic labour*. In this thesis, I argued that functional understanding of a technology is a prerequisite for meaningful ethical engagement, and that this understanding need not reside in the same person as the technical expertise that grounds it. This raises a deeper question: can technological understanding, as a cognitive skill, be shared across multiple people rather than residing in a single individual? The literature on distributed cognition and knowledge

in science (e.g. Giere, 2002) offers relevant resources, but whether understanding – as distinct from knowledge – can take a distributed form remains undertheorised.

This is a question that matters practically: if understanding can be distributed, it has implications for how we organise ethical engagement with technology, as well as for what we can reasonably expect from those involved. It may mean that meaningful ethical engagement does not require each participant to have full functional understanding, but rather that the group – e.g. a group tasked with assessing the ethical aspects of a technology – possesses sufficient understanding as a whole.

Related to this is the question of ethical expertise. The thesis argued that a functional understanding of technology is a necessary condition for meaningful ethical discussion, but not that it is *sufficient*. Meaningful ethical engagement also requires familiarity with ethical concepts, principles, and analytical methods – arguably, a form of ethical expertise. While ethicists certainly hold no 'moral monopoly' and are not the only ones authorised to reflect on what is right and wrong, they have developed a particular form of expertise in engaging with ethical issues. As Yoder puts it: "A claim to ethics expertise is not based on the truth of one's judgments but on one's ability to provide a coherent justification for them." (Yoder, 1998, p. 13). Such justification requires knowledge, skills, and a commitment to reasoned enquiry. These capacities can be cultivated and constitute a genuine form of expertise (Singer, 1972). While claims about ethical expertise seem to have lost favour in recent debates, the argument that technological understanding is a requirement for ethical discussion naturally raises the parallel question: What role does ethical expertise play in studying technology, and who is expected to have it?

5.2 Division of moral labour

The question about ethical expertise leads to a second avenue for future research: the *division of moral labour.* The ethical discussion of quantum technologies, as the chapters of Part II illustrated, involves many actors: scientists, engineers, companies, policymakers, ethicists, and many more. But who bears responsibility for engaging with the ethical dimensions of technological development? Rip (2017) uses the concept of the division of moral labour to describe how such responsibility is – or fails to be – distributed across actors in the innovation ecosystem. In practice, this division is often poorly organised or tacit, creating a situation of 'organised irresponsibility' in which everyone is nominally responsible while it remains unclear who is responsible for what (Swierstra & Rip, 2007). A better organised division of moral labour would involve explicit agreements about who bears responsibility at which stage of development, "distributing praise and blame before the fact" (Swierstra & Rip, 2007, p. 18). Shelley-Egan (2011; Rip & Shelley-Egan, 2019) raised this issue in the context of nanotechnology and called for opening up the division of roles and responsibilities to

explicit debate. That call is equally relevant in the context of quantum technologies. Without such reflection, ethical responsibilities risk remaining unassigned and unaddressed.

What has stood out to me, across all the discussions I have joined, is the relationship – or gap – between those who develop a technology ('the scientist') and those concerned with its societal implications ('the ethicist'). These actors do not always seem to understand or even appreciate one another – if they happen to meet at all. I have frequently spoken to scientists who draft their grant proposal ethics sections themselves, and experience this as 'ticking the box'. I have also participated in ethical discussions held without relevant technical expertise in the room. On the occasions when technical expertise was represented in the room, I have frequently sensed a dismissive attitude towards ethics, as if ethical reflection were an obstacle to innovation rather than a condition for responsible development. Questions from co-participants and students such as "What has ethics ever delivered?" and "Why would I bother innovating responsibly if others just do it fast?" are illustrative of this. It has often seemed to me that the 'responsible' and the 'innovation' of responsible innovation speak different languages, with too little genuine, mutually reflexive exchange between them.

Previous research on the division of moral labour has largely focused on its institutional dimensions, discussing how responsibilities are assigned across actors. What seems less explored, and what strikes me as equally important, is the cultural and epistemic dimension of this division: the different languages, norms, and ways of knowing that scientists and ethicists bring to the table, and the challenge of bridging them. Just as an ethicist cannot be expected to be a technical expert, a scientist or engineer cannot be expected to be an ethical expert. The two thus must join forces.

A division of moral labour is thus not only about dividing roles and responsibilities, but also about fostering collaboration between actors who bring complementary expertise. Lunshof and Rijssenbeek (2024) therefore call for a collaborative ethics: a closer collaboration between scientists and ethicists (see also Rip, 2009; Shelley-Egan, 2011). They propose a model for the ethical accompaniment of technoscientific research at different stages of development. Yet what fosters such collaboration in practice – the conditions that enable it to be genuinely productive – merits further investigation. It may involve embedding ethicists in research teams, developing shared vocabularies, and creating structured spaces for dialogue that make ethics more grounded in the technological realities and innovation more conscious of its normative dimensions.

These questions have not gone unnoticed in the literature. The technology ethics literature has engaged with the responsibilities of those involved in the development of technology and their enactment (e.g. Grunwald, 2022; Van de Poel & Royakkers, 2011). The adjacent fields of responsible research and innovation and ELSA – which focuses

on the ethical, legal, and social aspects of emerging science and technology – have also engaged with the roles and responsibilities of different types of stakeholders, mostly those developing technology (e.g. scientists, engineers) and those assessing it (e.g. social scientists/humanities scholars) (Rip, 2009; Calvert & Martin, 2009; Stegmaier, 2009; Shelley-Egan, 2011; Zwart et al., 2014; Pellé & Reber, 2015; Balmer et al., 2015). Notably, Shelley-Egan (2011) empirically investigated the perceived roles and responsibilities of scientists and industrial actors in the responsible development of nanotechnology, finding a largely traditional division of moral labour: science and industry developing technology, and others – social scientists, humanities scholars, societal actors – looking after the ethics. Based on my observations in the quantum context, this traditional division still appears to be in place. Future research should not only investigate whether this is indeed the case, but also examine why, more than a decade after this pattern was identified as a potential obstruction of responsible innovation, it persists.

This thesis motivates adding an epistemic angle to such research. The argument that meaningful ethical engagement requires a form of understanding raises a question that goes beyond who is and/or feels responsible for ethical engagement: who feels *epistemically equipped* to engage, and how that epistemic basis can be ensured. This connects to Stegmaier's observation that "shared concepts of collaboration need to be developed that function as common denominators" (2009, p. 117). He argues that productive collaboration between scientists and scholars working on the societal aspects of technology requires shared points of reference that make their different forms of expertise mutually intelligible. This thesis suggests that technological understanding can serve as one such common denominator: a shared grasp of what a technology is and may do provides the epistemic basis on which scientists and ethicists can meaningfully engage with each other's contributions and start a constructive conversation.

The analyses in Chapters 6–8 suggest that such a common denominator is currently lacking in the quantum discourse: claims about the abilities of future quantum technologies frequently exceed what current evidence supports, suggesting that functional understanding has not kept pace with – and in some cases has raced ahead of – technical understanding. Since functional understanding necessarily builds on technical understanding, and technical understanding of quantum technologies is itself still developing, this gap is to some extent unavoidable at the current stage of development. Rather than filling that gap with speculation, functional understanding should be advanced in tandem with technical progress. Recognising shared understanding as the common denominator for productive collaboration and grounded ethical debate reframes this as a joint responsibility that involves both scientists and ethicists.

Research into the epistemic conditions for collaboration could thus also shed new light on the roles and responsibilities discussion itself. If scientists do not see themselves as directly responsible – or equipped – for ethical engagement, they may nonetheless bear a responsibility for creating the conditions that make such engagement possible, for example by making their work intelligible to those assessing its implications. This reframes the question of roles and responsibilities from 'who is responsible for doing ethics' to 'who plays a role in enabling it'.

Moreover, a persisting traditional division of moral labour may be reinforced by the pronounced knowledge asymmetries between quantum scientists and ethicists. Attending to the epistemic dimension of ethical engagement could illuminate why collaboration remains difficult and, in turn, what it would take to make it genuinely productive. Part of that story is cultural: as long as the ethicist is seen as the party-pooper – an obstacle to innovation rather than a condition for responsible development – a genuinely collaborative division of moral labour cannot be realised. Empirically examining the perceived roles and responsibilities of scientists and ethicists in the quantum ecosystem – and what enables or obstructs genuine collaboration between them, with special attention to its epistemic conditions – would address a persistent challenge for responsible innovation, in the quantum domain and beyond.

6. Closing remarks

By illuminating what enables a meaningful ethical discussion of technology, this thesis aimed to contribute to the ethics of new and emerging technologies, and to advance the emerging ethical discussion about quantum technologies specifically. In doing so, it sought to promote reflexivity, asking, in Verhaeghe's words, "in which direction we are sliding", and whether we are right to call that progress. The thesis is a plea for both ethical ambition and epistemic humility in engaging with new and emerging technologies – consciously shaping the tools that will shape us.

References

Abdikhakimov, I. (2024). Preparing for a quantum future: Strategies for strengthening international data privacy in the face of evolving technologies. International Journal of Law and Policy, 2(5), 42–46.

AIVD, CWI & TNO (2024). PQC migration handbook: Guidelines for migrating to post-quantum cryptography (2nd ed.). https://english.aivd.nl/documents/2024/12/3/the-pqc-migration-handbook

Alleblas, J. (2026). Moralized visions and technological design [Doctoral dissertation]. ISBN/EAN: 978-94-6384-902-9

APRE & CDTI (n.d.). Guiding notes to use the TRL self-assessment tool. https://horizoneuropencpportal.eu/sites/default/files/2022-12/trl-assessment-tool-guide-final.pdf

Aslam, N., Zhou, H., Urbach, E. K., Turner, M. J., Walsworth, R. L., Lukin, M. D., & Park, H. (2023). Quantum sensors for biomedical applications. Nature Reviews Physics, 5(3), 157–169.

Atladóttir, D., Roch, N., Leese, M., et al. (2025). A quantum of responsibility? A comparison of national quantum governance frameworks and expert views. Digital Society, 4, 54. https://doi.org/10.1007/s44206-025-00205-7

Australian Government (2023). National quantum strategy: Building a thriving future with Australia's quantum advantage. https://industry.gov.au/quantum

Babbush, R., Zalcman, A., Gidney, C., Broughton, M., Khattar, T., Neven, H., Bergamaschi, T., Drake, J., & Boneh, D. (2026). Securing elliptic curve cryptocurrencies against quantum vulnerabilities: Resource estimates and mitigations. arXiv. https://arxiv.org/pdf/2603.28846

Balbus, J., Denison, R., Florini, K., & Walsh, S. (2005). Getting nanotechnology right the first time. Issues in Science and Technology, 21(4), 65–71.

Balmer, A., Calvert, J., Marris, C., Molyneux-Hodgson, S., Frow, E., Kearnes, M., Bulpin, K., Schyfter, P., Mackenzie, A., & Martin, P. (2015). Taking roles in interdisciplinary collaborations: Reflections on working in post-ELSI spaces in the UK synthetic biology community. Science and Technology Studies, 28(3), 3–25.

Barea, M. A. (2026). Quantum for whom? Situating ethical risks in quantum technologies' innovation context. Ethics and Society, 20(7). DOI: https://doi.org/10.1007/s11569-026-00496-1

Barman, K. G., Caron, S., Claassen, T., & De Regt, H. W. (2024). Towards a benchmark for scientific understanding in humans and machines. Minds and Machines, 34(1), 1–16. DOI: https://doi.org/10.1007/s11023-024-09657-1

Barnes, B. (1982). The science-technology relationship: A model and a query. Social Studies of Science, 12(1), 166–172.

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115. DOI: https://doi.org/10.1016/j.inffus.2019.12.012

Baumberger, C. (2019). Explicating objectual understanding: Taking degrees seriously. Journal for General Philosophy of Science, 50(3), 367–388.

Beige, A., Predojević, A., Metelmann, A., Sanpera, A., Macchiavello, C., Koch, C. P., & Ahufinger, V. (2025). Women for quantum – manifesto of values. Communications Physics, 8(1), 422.

Berling, T. V. (2011). Science and securitization: Objectivation, the authority of the speaker, and mobilization of scientific facts. Security Dialogue, 42, 385–397.

Bernstein, D., Hülsing, A., & Lange, T. (2022). Post-quantum cryptography: Integration study. European Union Agency for Cybersecurity. https://www.enisa.europa.eu/publications/post-quantum-cryptography-integration-study

Bernstein, M. J., Nielsen, M. W., Alnor, E., Brasil, A., Birkving, A. L., Chan, T. T., Griessler, E., de Jong, S., van de Klippe, W., Meijer, I., Yaghmaei, E., Busch Nicolaisen, P., Nieminen, M., Novitzky, P., & Mejlgaard, N. (2022). The societal readiness thinking tool: A practical resource for maturing the societal readiness of research projects. Science and Engineering Ethics, 28(1), 6.

Bijker, W., & Pinch, T. (1987). The social construction of facts and artifacts: Or how the sociology of science and the sociology of technology might benefit of each other. MIT Press.

Blok, V. (2021). What is innovation? Laying the ground for a philosophy of innovation. Techné: Research in Philosophy and Technology, 25(1), 72–96.

Blok, V., & Lemmens, P. (2015). The emerging concept of responsible innovation: Three reasons why it is questionable and calls for a radical transformation of

the concept of innovation. In Responsible innovation 2: Concepts, approaches, and applications (pp. 19–35). Springer.

Boenink, M., Swierstra, T., & Stemerding, D. (2010). Anticipating the interaction between technology and morality: A scenario study of experimenting with humans in bionanotechnology. Studies in Ethics, Law, and Technology, 4(2), Article 4. DOI: https://doi.org/10.2202/1941-6008.1098

Bontems, V. K. (2014). What does innovation stand for? Review of a watchword in research policies. Journal of Innovation Economics & Management, (3), 39–57.

Boon, M. (2006). How science is applied in technology. International Studies in the Philosophy of Science, 20(1), 27–47. DOI: https://doi.org/10.1080/02698590600640992

Boon, M. (2009). Understanding in the engineering sciences: Interpretative structures. In H. W. De Regt, S. Leonelli, & K. Eigner (Eds.), Scientific understanding: Philosophical perspectives (pp. 249–270). University Press.

Boon, M. (2015). The scientific use of technological instruments. In S. O. Hansson (Ed.), The role of technology in science: Philosophical perspectives (pp. 55–79). Springer Netherlands.

Boon, M. (2017). Measurements in the engineering sciences: An epistemology of producing knowledge of physical phenomena. In N. Mößner & A. Nordmann (Eds.), Reasoning in measurement (pp. 203–219). Routledge.

Boon, M. (2020). Scientific methodology in the engineering sciences. In D. Michelfelder & N. Doorn (Eds.), The Routledge handbook of the philosophy of engineering (pp. 80–94). Routledge.

Borup, M., Brown, N., Konrad, K., & Van Lente, H. (2006). The sociology of expectations in science and technology. Technology Analysis & Strategic Management, 18(3–4), 285–298.

Boulet, C. A., Kaya, O., MacKinnon, P., & Rowell, M. (2021). Quantum-safe cybersecurity talent and job market analysis (2020–21). Quantum-Safe Canada.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. DOI: https://doi.org/10.1191/1478088706qp063oa

Brey, P. (2012). Anticipatory ethics for emerging technologies. NanoEthics, 6(1), 1–13.

Brey, P. (2017). Ethics of emerging technologies. In S. O. Hansson (Ed.), Methods for the ethics of technology (pp. 175–192). Rowman and Littlefield International.

Broussard, M. (2019). Artificial unintelligence: How computers misunderstand the world. MIT Press. DOI: https://doi.org/10.7551/mitpress/11022.001.0001

Bruno, I., Lobo, G., Covino, B. V., Donarelli, A., Marchetti, V., Panni, A. S., & Molinari, F. (2020). Technology readiness revisited: A proposal for extending the scope of impact assessment of European public services. In Proceedings of the 13th international conference on theory and practice of electronic governance (pp. 369–380).

Buchanan, W., & Woodward, A. (2017). Will quantum computers be the end of public key encryption? Journal of Cyber Security Technology, 1(1), 1–22.

Burgess, A. (2015). Social construction of risk. In H. Cho, T. Reimer, & K. A. McComas (Eds.), The SAGE handbook of risk communication (pp. 56–68). SAGE.

Cain, M., Xu, Q., King, R., Picard, L. R. B., & Levine, H. (2026). Shor's algorithm is possible with as few as 10,000 reconfigurable atomic qubits. arXiv. https://arxiv.org/pdf/2603.28627

Calvert, J., & Martin, P. (2009). The role of social scientists in synthetic biology. Science & Society Series on Convergence Research. EMBO Reports, 10(3), 201.

Cao, Y., Romero, J., & Aspuru-Guzik, A. (2018). Potential of quantum computing for drug discovery. IBM Journal of Research and Development, 62(6), 6:1–6:20. DOI: https://doi.org/10.1147/JRD.2018.2888987

Capraro, V., Lentsch, A., Acemoglu, D., Akgun, S., Akhmedova, A., Bilancini, E., & Viale, R. (2024). The impact of generative artificial intelligence on socioeconomic inequalities and policy making. PNAS Nexus, 3(6), p.191.

Carroll, J. (2004). Completing design in use: Closing the appropriation cycle. In Proceedings of the ECIS, 44.

Chen, L., Jordan, S., Liu, Y.-K., Moody, D., Peralta, R., Perlner, R., & Smith-Tone, D. (2016). Report on post-quantum cryptography (NIST Internal Report 8105). National Institute of Standards and Technology. DOI: https://doi.org/10.6028/NIST.IR.8105

Chui, M., Manyika, J., Miremadi, M., Henke, N., Chung, R., Nel, P., & Malhotra, S. (2018). Notes from the AI frontier: Insights from hundreds of use cases (pp. 1–31). McKinsey Global Institute.

Coeckelbergh, M., & Sætra, H. S. (2023). Climate change and the political pathways of AI: The technocracy-democracy dilemma in light of artificial intelligence and human agency. Technology in Society, 75, Article 102406.

Coenen, C., & Grunwald, A. (2017). Responsible research and innovation (RRI) in quantum technology. Ethics and Information Technology, 19(4), 277–294.

Collingridge, D. (1980). The social control of technology. St. Martin's Press.

Culkin, J. M. (1967, March 18). A schoolman's guide to Marshall McLuhan. Saturday Review, 51–53, 71–72.

Danaher, J. (2022). Techno-optimism: An analysis, an evaluation and a modest defence. Philosophy & Technology, 35(2), 54.

Danisch, R. (2010). Political rhetoric in a world risk society. Rhetoric Society Quarterly, 40, 172–192.

Danisch, R. (2011). Risk assessment as rhetorical practice: The ironic mathematics behind terrorism, banking, and public policy. Public Understanding of Science, 21, 1–16.

De Haro, S. (in preparation). A theory of technological understanding and technology assessment.

De Haro, S., & Butterfield, J. (2025). The philosophy and physics of duality. Oxford University Press.

De Haro, S., & De Regt, H. W. (2020). A precipice below which lies absurdity? Theories without a spacetime and scientific understanding. Synthese, 197, 3121–3149.

De Jong, E. (2022). Own the unknown: An anticipatory approach to prepare society for the quantum age. Digital Society, 1(2), 15.

De Jong, E. (2025). Ethics readiness of technology: The case for aligning ethical approaches with technological maturity. *Science and Engineering Ethics, 31*(5), 30. DOI: https://doi.org/10.1007/s11948-025-00556-x

De Jong, E., De Haro, S. (2026). Technological understanding. On the cognitive skill involved in the design and use of technological artefacts. *Synthese, 207*(121). DOI: https://doi.org/10.1007/s11229-026-05503-2

De Jong, E. (2026). Functional understanding of quantum technologies is essential to the ethical debate about their impact. *Ethics and Society, 20*(21). DOI: https://doi.org/10.1007/s11569-026-00501-7

De Jong, E., & De Haro, S. (2026). A contextual approach to technological understanding and its assessment. [preprint]. https://arxiv.org/abs/2503.21437

De Jong, E., & De Haro, S. (2026). Better secure than sorry? Precaution in the quantum-safe cryptography transition. [under review]. In: *Ethics and Information Technology.*

De Jong, E., & Grunwald, A., (2026). Visions of Quantum Computing: An Ethical Assessment. *Futures*, 103941. DOI: https://doi.org/10.1016/j.futures.2026.103941

De Jong, E., & Shelley-Egan, C. (2026). Quantum for Good through a lens of technosolutionism: A critical-constructive view. *Journal of Responsible Technology, 25*, 100157. DOI: https://doi.org/10.1016/j.jrt.2026.100157

De Regt, H. W. (2009). The epistemic value of understanding. Philosophy of Science, 76(5), 585–597. DOI: https://doi.org/10.1086/605795

De Regt, H. W. (2013). Understanding and explanation: Living apart together? Studies in History and Philosophy of Science Part A, 44(3), 505–509. DOI: https://doi.org/10.1016/j.shpsa.2012.12.002

De Regt, H. W. (2015). Scientific understanding: Truth or dare? Synthese, 192(12), 3781–3797.

De Regt, H. W. (2017). Understanding scientific understanding. Oxford University Press.

De Regt, H. W., & Dieks, D. (2005). A contextual approach to understanding. Synthese, 144, 137–170.

De Wolf, R. (2017). The potential impact of quantum computers on society. Ethics and Information Technology, 19(4), 271–276. DOI: https://doi.org/10.1007/s10676-017-9439-z

Dignum, M., Correljé, A., Groenleer, M., & Scholten, D. (2018). Governing through visions: Evaluating the performativity of the European gas target models. Energy Research & Social Science, 35, 193–204.

Doyle, L., Seifollahi, F., & Singh, C. (2026). Do we have a quantum computer? Expert perspectives on current state and future prospects. Physical Review Physics Education Research, 22(1), 010101.

Drougkas, A., Komzaite, U., Philippou, E., Abel, P., Gratiolet, F., & Maaskant, E. (2024). NIS investments 2024: Cybersecurity policy assessment. European Union Agency for Cybersecurity. https://www.enisa.europa.eu/publications/nis-investments-2024

Dyakonov, M. (2019). When will useful quantum computers be constructed? IEEE Spectrum, 56(3), 24–29.

European Commission, NIS Cooperation Group – PQC Workstream (2025). A coordinated implementation roadmap for the transition to post-quantum cryptography: Part 1 (Version 1.1). https://digital-

strategy.ec.europa.eu/en/library/coordinated-implementation-roadmap-transition-post-quantum-cryptography

Executive Office of the President of the United States (2024, July). Report on post-quantum cryptography as required by the Quantum Computing Cybersecurity Preparedness Act, Public Law No. 117-260. https://bidenwhitehouse.archives.gov/wp-content/uploads/2024/07/REF_PQC-Report_FINAL_Send.pdf

Ezratty, O. (2022). Mitigating the quantum hype. arXiv preprint arXiv:2202.01925.

Ferrari, A. (2010). Developments in the debate on nanoethics: Traditional approaches and the need for new kinds of analysis. NanoEthics, 4(1), 27–52.

Ferrari, A. (2013). From nanoethics to the normativity of technological visions. Etica & Politica, 15(1), 220–235.

Ferrari, A., Coenen, C., & Grunwald, A. (2012). Visions and ethics in current discourse on human enhancement. NanoEthics, 6(3), 215–229. DOI: https://doi.org/10.1007/s11569-012-0155-1

Floridi, L., & Strait, A. (2020). Ethical foresight analysis: What it is and why it is needed? The 2020 Yearbook of the Digital Ethics Lab, 173–194.

Friedman, M. (1974). Explanation and scientific understanding. The Journal of Philosophy, 71(1), 5–19.

Frigg, R. (2023). Models and theories. Routledge.

Gardiner, S. M. (2006). A core precautionary principle. The Journal of Political Philosophy, 14(1), 33–60.

Gasser, U., De Jong, E., & Kop, M. (2024). A call for responsible quantum technology. Nature Physics, 20(4), 525–527.

Genus, A., & Stirling, A. (2018). Collingridge and the dilemma of control: Towards responsible and accountable innovation. Research Policy, 47(1), 61–69.

Gercek, A. A., & Seskir, Z. C. (2025). Navigating the quantum divide(s). IEEE Engineering Management Review, 53(6), 168–193. DOI: https://doi.org/10.1109/EMR.2025.3547927

Gidney, C. (2025). How to factor 2048 bit RSA integers with less than a million noisy qubits. arXiv. https://arxiv.org/abs/2505.15917

Gidney, C., & Ekerå, M. (2021). How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits. Quantum, 5, 433.

Giere, R. (2002). Scientific cognition as distributed cognition. In P. Carruthers, S. Stitch & M. Siegal (Eds.), Cognitive bases of science (pp. 285–299). Cambridge University Press.

Gilbert, F., & Goddard, E. (2014). Thinking ahead too much: Speculative ethics and implantable brain devices. AJOB Neuroscience, 5(1), 49–51.

Gisin, N., & Thew, R. (2007). Quantum communication. Nature Photonics, 1(3), 165–171.

Goodman, B., & Flaxman, S. (2017). European union regulations on algorithmic decision-making and a "right to explanation". AI Magazine, 38(3), 50–57.

Grimm, S. (2012). The value of understanding. Philosophy Compass, 7(2), 103–117.

Grimm, S. R. (2010). The goal of explanation. Studies in History and Philosophy of Science, 41, 337–344. DOI: https://doi.org/10.1016/j.shpsa.2010.10.006

Grinbaum, A. (2017). Narratives of quantum theory in the age of quantum technologies. Ethics and Information Technology, 19(4), 295–306.

Grinbaum, A., & Groves, C. (2013). What is "responsible" about responsible innovation? Understanding the ethical issues. In Responsible innovation: Managing the responsible emergence of science and innovation in society (pp. 119–142). Wiley.

Grunwald, A. (2004). Vision assessment as a new element of the technology futures analysis toolbox. In *Proceedings of the EU-US Scientific Seminar: New Technology Foresight, Forecasting & Assessment Methods, Seville,* May 13–14, 2004.

Grunwald, A. (2005). Nanotechnology: A new field of ethical inquiry? *Science & Engineering Ethics, 11*(2), 187–201. DOI: https://doi.org/10.1007/s11948-005-0041-0

Grunwald, A. (2007). Converging technologies: Visions, increased contingencies of the conditio humana, and search for orientation. *Futures*, *39*(4), 380-392.

Grunwald, A. (2010). From speculative nanoethics to explorative philosophy of nanotechnology. *NanoEthics, 4*, 91–101.

Grunwald, A. (2011). Responsible innovation: Bringing together technology assessment, applied ethics, and STS research. *Enterprise and Work Innovation Studies, 7*, 9–31.

Grunwald, A. (2014). The hermeneutic side of responsible research and innovation. Journal of Responsible Innovation, 1(3), 274–291.

Grunwald, A. (2017). Assigning meaning to NEST by technology futures: Extended responsibility of technology assessment in RRI. Journal of Responsible Innovation, 4(2), 100–117.

Grunwald, A. (2018). Technology assessment in practice and theory. Routledge.

Grunwald, A. (2020). The objects of technology assessment: Hermeneutic extension of consequentialist reasoning. Journal of Responsible Innovation, 7(1), 96–112. DOI: https://doi.org/10.1080/23299460.2019.1647086

Grunwald, A. (2022). The responsibility of researchers and engineers: Codes of ethics for emerging technologies. In Codes of ethics and ethical guidelines: Emerging technologies, changing fields (pp. 243–258). Springer International Publishing.

Grunwald, A. (2023). The spreading of techno-visionary futures. In Diffusive spreading in nature, technology and society (pp. 309–323). Springer International Publishing.

Grunwald, A. (2026). Technology assessment for a 'good' Anthropocene. In Handbook of technology assessment (pp. 441–450). Edward Elgar Publishing.

Guston, D. H., & Sarewitz, D. (2020). Real-time technology assessment. Technology in Society, 24, 93–109.

Hansen, L., & Nissenbaum, H. (2009). Digital disaster, cyber security, and the Copenhagen School. International Studies Quarterly, 53(4), 1155–1175.

Hansson, S. O. (2003). Ethical criteria of risk acceptance. Erkenntnis, 59(3), 291–309.

Hansson, S. O. (Ed.). (2017). The ethics of technology: Methods and approaches. Bloomsbury Publishing USA.

Hansson, S. O. (2010). Risk: Objective or subjective, facts or values. Journal of Risk Research, 13(2), 231–238.

Hansson, S. O. (2023). Risk. In E. N. Zalta & U. Nodelman (Eds.), The Stanford Encyclopedia of Philosophy (Summer 2023 Edition). https://plato.stanford.edu/archives/sum2023/entries/risk/

Hasan, K. F., Simpson, L., Baee, M. A. R., Islam, C., Rahman, Z., Armstrong, W., Gauravaram, P., & McKague, M. (2024). A framework for migrating to post-quantum cryptography: Security dependency analysis and case studies. IEEE Access, 12, 23427–23450. DOI: https://doi.org/10.1109/ACCESS.2024.3360412

Hempel, C. G. (1965). Scientific explanation: Essays in the philosophy of science. The Free Press.

Hidary, J. D. (2019). A brief history of quantum computing. In Quantum computing: An applied approach. Springer. DOI: https://doi.org/10.1007/978-3-030-23922-0_2

Hills, A. (2016). Understanding why. Noûs, 50(4), 661–688.

Hoofnagle, C. J., & Garfinkel, S. L. (2022). Law and policy for the quantum age. Cambridge University Press.

Houkes, W. (2009). The nature of technological knowledge. In A. Meijers (Ed.), Philosophy of technology and engineering sciences (pp. 309–350). Elsevier.

Houkes, W., & Meijers, A. (2022). Engineering knowledge. In S. Vallor (Ed.), The Oxford handbook of the philosophy of technology (pp. 128–153). Oxford University Press.

Houkes, W., & Vermaas, P. E. (2010). Technical functions: On the use and design of artefacts. Springer.

Hughes, J. (2009). An artifact is to use: An introduction to instrumental functions. Synthese, 168, 179–199. DOI: https://doi.org/10.1007/s11229-008-9335-5

Inglesant, P., Jirotka, M., & Hartswood, M. (2018). Responsible innovation in quantum technologies applied to defence and national security. NQIT.

Inglesant, P., Ten Holter, C., Jirotka, M., & Williams, R. (2021). Asleep at the wheel? Responsible innovation in quantum computing. Technology Analysis & Strategic Management, 33(11), 1364–1376.

ISACA (2025). Despite rising concerns, 95% of organizations lack a quantum computing roadmap, ISACA finds [Press release]. https://www.isaca.org/about-us/newsroom/press-releases/2025/organizations-lack-a-quantum-computing-roadmap-isaca-finds

Jacobson, I. (2004). Use cases — Yesterday, today, and tomorrow. Software & Systems Modeling, 3(3), 210–220.

Jaschke, D., & Montangero, S. (2023). Is quantum computing green? An estimate for an energy-efficiency quantum advantage. Quantum Science and Technology, 8(2), 025001.

Jasanoff, S. (2020). Imagined worlds: The politics of future-making in the twenty-first century. In The politics and science of prevision (pp. 27–44). Routledge.

Jensen, R. E. (2015). Rhetoric of risk. In H. Cho, T. Reimer, & K. A. McComas (Eds.), The SAGE handbook of risk communication (pp. 86–98). SAGE.

Johnston, S. (2018). Alvin Weinberg and the promotion of the technological fix. Technology and Culture, 59(3), 620–651.

Joly, P. B. (2015). Governing emerging technologies? The need to think outside the (black) box. In Science and democracy (pp. 133–155). Routledge.

Joseph, D., Misoczki, R., Manzano, M., Tricot, J., Dominguez Pinuaga, F., Lacombe, O., Leichenauer, S., Hidary, J., Venables, P., & Hansen, R. (2022). Transitioning organizations to post-quantum cryptography. Nature, 605, 237–243. DOI: https://doi.org/10.1038/s41586-022-04623-2

Kelp, C. (2015). Understanding phenomena. Synthese, 192(12), 3799–3816.

Kiran, A. H., Oudshoorn, N., & Verbeek, P. P. (2015). Beyond checklists: Toward an ethical-constructive technology assessment. Journal of Responsible Innovation, 2(1), 5–19.

Kitcher, P. (1981). Explanatory unification. Philosophy of Science, 48, 507–531.

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher & W. Salmon (Eds.), Scientific explanation (pp. 410–505). University of Minnesota Press.

Königs, P. (2022). What is techno-optimism? Philosophy & Technology, 35(3), 63.

Kong, I., Janssen, M., & Bharosa, N. (2022). Challenges in the transition towards a quantum-safe government. In Proceedings of the 23rd Annual International Conference on Digital Government Research (pp. 282–292).

Kop, M. (2023). Quantum-ELSPI: A novel field of research. Digital Society, 2(2), 20.

Kop, M., Aboy, M., De Jong, E., Gasser, U., Minssen, T., Cohen, I. G., Brongersma, M., Quintel, T., Floridi, L., & Laflamme, R. (2024a). Towards responsible quantum technology: Safeguarding, engaging and advancing quantum R&D. UC Law Science and Technology Journal, 15(1), 63.

Kop, M., Aboy, M., De Jong, E., Gasser, U., Minssen, T., Cohen, I. G., Brongersma, M., Quintel, T., Floridi, L., & Laflamme, R. (2024b). Ten principles for responsible quantum innovation. Quantum Science and Technology, 9(3), 035013.

Krelina, M. (2021). Quantum technology for military applications. EPJ Quantum Technology, 8(1), 24.

Krenn, M., Pollice, R., Guo, S. Y., Aldeghi, M., Cervera-Lierta, A., Friederich, P., dos Passos Gomes, G., Häse, F., Jinich, A., Nigam, A., Yao, Z., & Aspuru-Guzik, A. (2022). On scientific understanding with artificial intelligence. Nature Reviews Physics, 4(12), 761–769. DOI: https://doi.org/10.1038/s42254-022-00518-3

Kroes, P. (2002). Design methodology and the nature of technical artefacts. Design Studies, 23(3), 287–302. DOI: https://doi.org/10.1016/S0142-694X(01)00039-4

Kudina, O. (2021). "Alexa, who am I?": Voice assistants and hermeneutic lemniscate as the technologically mediated sense-making. Human Studies, 44(2), 233–253.

Kudina, O., & Verbeek, P. P. (2019). Ethics from within: Google Glass, the Collingridge dilemma, and the mediated value of privacy. Science, Technology, & Human Values, 44(2), 291–314.

Kuorikoski, J., & Ylikoski, P. (2015). External representations and scientific understanding. Synthese, 192, 3817–3837.

Kvanvig, J. L. (2003). The value of knowledge and the pursuit of understanding. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511498909

Latour, B. (2011). Love your monsters: Why we must care for our technologies as we do our children. Breakthrough Journal, Fall, 19–26.

Le, T. D., Do, P. H., Dinh, T. D., & Pham, V. D. (2025). Are enterprises ready for quantum-safe cybersecurity? arXiv. https://arxiv.org/abs/2509.01731

LeCain, T. L. (2004). When everybody wins does the environment lose? The environmental techno-fix in twentieth-century mining. In L. Rosner (Ed.), The technological fix: How people use technology to create and solve social problems (pp. 137–153). Routledge.

Lin, J., Ji, Y., Zhao, J., Tong, X., & Yi, X. (2022). Development strategy of quantum-based deep geophysical exploration technology and equipment. Strategic Study of Chinese Academy of Engineering, 24(4), 156.

Lindsay, J. R. (2020). Surviving the quantum cryptocalypse. Strategic Studies Quarterly, 14(2), 49–73. https://www.jstor.org/stable/26915277

Lipton, P. (2004). Inference to the best explanation. Routledge.

Loebenberger, D., Hirsch, E., Gazdag, S.-L., Herzinger, D., Näther, C., & Steghöfer, J.-P. (2024). On the formalization of cryptographic migration. arXiv. https://arxiv.org/abs/2408.05997

Lucivero, F., Swierstra, T., & Boenink, M. (2011). Assessing expectations: Towards a toolbox for an ethics of emerging technologies. NanoEthics, 5, 129–141.

Lunshof, J. E., & Rijssenbeek, J. (2024). Collaborative ethics: Innovating collaboration between ethicists and life scientists. Nature Methods, 21(9), 1571–1574.

Luo, W., Cao, L., Shi, Y., Wan, L., Zhang, H., Li, S., Chen, G., Li, Y., Li, S., Wang, Y., Sun, S., Karim, M. F., Cai, H., Kwek, L. C., & Liu, A. Q. (2023). Recent progress in quantum photonic chips for quantum communication and internet. Light: Science & Applications, 12, 175. DOI: https://doi.org/10.1038/s41377-023-01173-8

Malina, L., Dzurenda, P., Ricci, S., Hajny, J., Srivastava, G., Matulevičius, R., & Tang, Q. (2021). Post-quantum era privacy protection for intelligent infrastructures. IEEE Access, 9, 36038–36077.

Manders-Huits, N. (2011). What values in design? The challenge of incorporating moral values into design. Science and Engineering Ethics, 17(2), 271–287.

Manson, N. A. (2012). Formulating the precautionary principle. In A. Bird & J. Ladyman (Eds.), Arguing about science (pp. 607–616). Routledge.

Mehnert, W. (2025). Beyond the bad actor narrative: Reconsidering ethical imagination in quantum computing. NanoEthics, 19(3), 21.

Mehnert, W., & Grunwald, A. (2024). Hermeneutic technology assessment. In Handbook of technology assessment (pp. 281–290). Edward Elgar Publishing.

Meier, F., & Yamasaki, H. (2025). Energy-consumption advantage of quantum computation. PRX Energy, 4(2), 023008.

Meinsma, A. L., Albers, C. J., Vermaas, P., Smeets, I., & Cramer, J. (2024). The effect of frames on engagement with quantum technology. arXiv preprint. https://arxiv.org/abs/2404.14104

Meinsma, A. L., Kristensen, S. W., Reijnierse, W. G., Smeets, I., & Cramer, J. (2023). Is everything quantum 'spooky and weird'? An exploration of popular communication about quantum science and technology in TEDx talks. Quantum Science and Technology, 8(3), 035004.

Meinsma, A., Rothe, T., Reijnierse, W., Smeets, I., & Cramer, J. (2025). Quantum in the media: A content analysis of Dutch newspapers. Science Communication, 48(1), 128–147.

Michael, M. (2000). Futures of the present: From performativity to prehension. In N. Brown, B. Rappert, & A. Webster (Eds.), Contested futures: A sociology of prospective techno-science. Ashgate.

Miller, S. (2022). Epistemic actions, abilities and knowing-how: A non-reductive account. Social Epistemology, 36(4), 466–485. DOI: https://doi.org/10.1080/02691728.2021.2020931

Möller, M., & Vuik, C. (2017). On the impact of quantum computing technology on future developments in high-performance scientific computing. Ethics and Information Technology, 19(4), 253–269.

Morgan, M., & Morrison, M. (1999). Models as mediators: Perspectives on natural and social science. Cambridge University Press.

Morozov, E. (2013). To save everything click here: Technology, solutionism, and the urge to fix problems that don't exist. PublicAffairs.

Mosca, M., & Piani, M. (2026). Quantum threat timeline report 2025. Global Risk Institute. https://globalriskinstitute.org/publication/quantum-threat-timeline-report-2025b

Munthe, C. (2011). The price of precaution and the ethics of risk. Springer.

Nachtwey, O., & Seidl, T. (2024). The solutionist ethic and the spirit of digital capitalism. Theory, Culture & Society, 41(2), 91–112.

Naess, A. (1973). The shallow and the deep, long-range ecology movement: A summary. Inquiry, 16, 95–100.

Nahar, S., Pithawa, D., Bhardwaj, V., Rawat, R., Rawat, A., & Pachlasiya, K. (2023). Quantum technology for military applications. In Quantum computing in cybersecurity (pp. 313–334).

National Academies of Sciences, Engineering, and Medicine (2019). Quantum computing: Progress and prospects. The National Academies.

National Institute of Standards and Technology (2024, August 13). NIST releases first 3 finalized post-quantum encryption standards. https://www.nist.gov/news-events/news/2024/08/nist-releases-first-3-finalized-post-quantum-encryption-standards

Netherlands Enterprise Agency (2022). Technology readiness levels. https://www.rvo.nl/onderwerpen/trl

Nguyen, A. (2025). Export controls as innovation marketing? Sociotechnical imaginaries in the ringfencing of quantum technologies. Law, Technology and Humans, 7(1), 68–83.

NIST (National Institute of Standards and Technology) (2024, 12 November). Transition to post-quantum cryptography standards [Web page]. https://csrc.nist.gov/pubs/ir/8547/ipd

Nixdorff, K., Borisova, T., Komisarenko, S., & Dando, M. (2018). Dual-use nano-neurotechnology: An assessment of the implications of trends in science and technology. Politics and the Life Sciences, 37(2), 180–202.

Nordmann, A. (2007). If and then: A critique of speculative nanoethics. NanoEthics, 1(1), 31–46.

Nordmann, A. (2014). Responsible innovation, the art and craft of anticipation. Journal of Responsible Innovation, 1(1), 87–98.

Nordmann, A., & Grunwald, A. (2023). Hermeneutic technology assessment — Why it is needed and what it might be. In A. Grunwald, A. Nordmann, & M. Sand (Eds.), Hermeneutics, history, and technology: The call of the future (pp. 37–41). Routledge.

Nordmann, A., & Rip, A. (2009). Mind the gap revisited. Nature Nanotechnology, 4(5), 273–274.

Nydal, R., De Grandis, G., & Ursin, L. (2023). When is a techno-fix legitimate? The case of viticultural climate resilience. Journal of Agricultural and Environmental Ethics, 36(1), 3.

O'Neil, C. (2017). Weapons of math destruction: How big data increases inequality and threatens democracy. Penguin Books.

OECD (2005). Oslo manual: Guidelines for collecting and interpreting innovation data (3rd ed.). OECD & Statistical Office of the European Communities.

Open Quantum Institute (n.d.). The Open Quantum Institute [Web page]. https://open-quantum-institute.cern/

Open Quantum Institute (2024). White paper 2024: SDG use cases. https://open-quantum-institute.cern/wp-content/uploads/2024/12/OQI_WhitePaper2024.pdf

Ott, D., & Peikert, C. (2019). Identifying research challenges in post quantum cryptography migration and cryptographic agility. arXiv. https://arxiv.org/abs/1909.07353

Owen, R., Stilgoe, J., Macnaghten, P., Gorman, M., Fisher, E., & Guston, D. (2013). A framework for responsible innovation. In R. Owen, J. Bessant, & M. Heintz (Eds.), Responsible innovation: Managing the responsible emergence of science and innovation in society (pp. 27–50). Wiley.

Palm, E., & Hansson, S. O. (2006). The case for ethical technology assessment (eTA). Technological Forecasting and Social Change, 73(5), 543–558.

Paudel, H. P., Syamlal, M., Crawford, S. E., Lee, Y.-L., Shugayev, R. A., Lu, P., Ohodnicki, P. R., Mollot, D., & Duan, Y. (2022). Quantum computing and simulations for energy applications: Review and perspective. ACS Engineering Au, 2(3), 151–196. DOI: https://doi.org/10.1021/acsengineeringau.1c00033

Pei, H., Fan, W., Duan, L., Liu, F., Pang, H., Li, R., Feng, Y., Liu, Y., Wang, Y., Yuan, H., Tang, J., Zheng, H., Qin, J., & Quan, W. (2025). Navigation in the future: Review of quantum sensing in navigation. Science China Physics, Mechanics & Astronomy, 68, 290301. DOI: https://doi.org/10.1007/s11433-025-2699-1

Pellé, S., & Reber, B. (2015). Responsible innovation in the light of moral responsibility. Journal on Chain and Network Science, 15(2), 107–118.

Perrier, E. (2021). Ethical quantum computing: A roadmap. arXiv preprint. https://arxiv.org/pdf/2102.00759

Phillipson, F. (2024). Quantum computing in logistics and supply chain management: An overview. arXiv preprint arXiv:2402.17520.

Possati, L. M. (2023). Ethics of quantum computing: An outline. Philosophy & Technology, 36(3), 48.

Possati, L. M. (2024). Quantum technologies: A hermeneutic technology assessment approach. NanoEthics, 18(1), 2.

Preskill, J. (1998). Quantum computing: Pro and con. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 454(1969), 469–486.

Preskill, J. (2018). Quantum computing in the NISQ era and beyond. arXiv. https://arxiv.org/pdf/1801.00862

Price, H. (2024). Risk and scientific reputation: Lessons from cold fusion. In C. Rhodes (Ed.), Managing extreme technological risk (pp. 25–85). World Scientific Publishing Europe.

Propp, T. B., Todd, B., Metwalli, S. A., Gallardo, A. H. S., Dascal, M., Ruffner, D., & Sanmartí-Vila, L. (2026). Meeting the needs of the global quantum science community: A call to action. arXiv preprint arXiv:2601.14297.

Pupillo, L., Ferreira, A., Lipiainien, V., & Polito, C. (2023). Quantum technologies and cybersecurity: Technology, governance and policy challenges. Centre for European Policy Studies.

Purohit, M., Chauhan, R., & Rawat, R. (2024). A cryptographic analysis of quantum computing. In V. Sharmila et al. (Eds.), Challenges in information, communication and computing technology (pp. 139–144). CRC Press.

Pyrrho, M. (2011). Nanotechnology and ethics: Assessing the unforeseeable. In Bioethics in the 21st century (pp. 121–138). Intech.

Quantum 2025 (n.d.). [Web page]. https://quantum2025.org/

Quantum Computing Cybersecurity Preparedness Act, H.R. 7535, 117th Cong., Pub. L. No. 117-260 (2022). https://www.congress.gov/bill/117th-congress/house-bill/7535

Quantum Delta Nederland (2019). Nationale Agenda Quantumtechnologie. https://open.overheid.nl/documenten/ronl-61f96d64-5471-48e7-aaeb-aad21a953d2c/pdf

QDNL (2024, 23 September). New quantum for good partnership initiative between UNICC and Quantum Delta NL [Web page]. https://quantumdelta.nl/news/new-quantum-for-good-partnership-initiative-between-unicc-and-quantum-delta-nl

Quantum for Good (n.d.). Leveraging quantum for good: A global collaborative movement [Web page]. https://quantumforgood.eu/#introduction

Rathenau Instituut (2023). Rathenau Scan: Quantumtechnologie in de samenleving. https://www.rathenau.nl/nl/digitalisering/rathenau-scan-quantumtechnologie-de-samenleving

Rawls, J. (1999). *A Theory of Justice*. (Revised Edition 1999). Cambridge, MA: Harvard University Press.

Reijers, W., Wright, D., Brey, P., Weber, K., Rodrigues, R., O'Sullivan, D., & Gordijn, B. (2018). Methods for practising ethics in research and innovation: A literature review, critical analysis and recommendations. Science and Engineering Ethics, 24(5), 1437–1481.

Reutlinger, A., Hangleiter, D., & Hartmann, S. (2018). Understanding (with) toy models. The British Journal for the Philosophy of Science, 69(4), 1069–1099.

Rip, A. (2009). Futures of ELSA. Science & Society Series on Convergence Research. EMBO Reports, 10(7), 666.

Rip, A. (2017). Division of moral labour as an element in the governance of emerging technologies. In Embedding new technologies into society (pp. 115–129). Jenny Stanford Publishing.

Rip, A. (2018). The past and future of RRI. In Futures of science and technology in society (pp. 115–133). Springer Fachmedien Wiesbaden.

Rip, A., & Kulve, H. T. (2008). Constructive technology assessment and socio-technical scenarios. In E. Fisher, C. Selin, & J. M. Wetmore (Eds.), Presenting futures. The yearbook of nanotechnology in society (Vol. 1, pp. 49–70). Springer. DOI: https://doi.org/10.1007/978-1-4020-8416-4_4

Rip, A., & Shelley-Egan, C. (2019). Positions and responsibilities in the 'real' world of nanotechnology. In Nanotechnology and its governance (pp. 101–107). Routledge.

Roache, R. (2008). Ethics, speculation, and values. NanoEthics, 2, 317–327.

Roberson, T. (2023). Talking about responsible quantum: "Awareness is the absolute minimum that... we need to do." NanoEthics, 17(1), 2.

Roberson, T. M. (2021). On the social shaping of quantum technologies: An analysis of emerging expectations through grant proposals from 2002–2020. Minerva, 59(3), 379–397.

Roberson, T., Leach, J., & Raman, S. (2021). Talking about public good for the second quantum revolution: Analysing quantum technology narratives in the

context of national strategies. Quantum Science and Technology, 6(2), Article 025001.

Robertson, A., Brooke, S., De Haro, S., & Schaffner, C. (2025). Irrelevant carrots and non-existent sticks: Trust, governance, and security in the transition to quantum-safe systems. arXiv. https://arxiv.org/abs/2512.16974

Root, D. (2025). Quantum technologies in the context of climate change: Emphasizing sustainability in a responsible innovation approach to quantum innovation. NanoEthics, 19(1), 4.

Rotolo, D., Hicks, D., & Martin, B. R. (2015). What is an emerging technology? Research Policy, 44(10), 1827–1843.

Ruser, A., & Machin, A. (2016). Technology can save us, can't it? The emergence of the 'technofix' narrative in climate politics. In: Proceedings of the international conference "Technology + Society = Future", pp. 437–447.

Ryan, M., & Blok, V. (2023). Stop re-inventing the wheel: Or how ELSA and RRI can align. Journal of Responsible Innovation, 10(1), 2196151.

Ryan, M., De Roo, N., Wang, H., Blok, V., & Atik, C. (2024). AI through the looking glass: An empirical study of structural social and ethical challenges in AI. AI & Society, 1–17.

Sætra, H. S. (2023). Introduction: The promise and pitfalls of techno-solutionism. In Technology and sustainable development (pp. 1–9). Routledge.

Sætra, H. S., & Danaher, J. (2022). To each technology its own ethics: The problem of ethical proliferation. Philosophy & Technology, 35(4), 93.

Sætra, H. S., & Selinger, E. (2024). Technological remedies for social problems: Defining and demarcating techno-fixes and techno-solutionism. Science and Engineering Ethics, 30(6), 1–17. DOI: https://doi.org/10.1007/s11948-024-00524-x

Sainz, A. B. (2022). Q-turn: Changing paradigms in quantum science. Quantum Science & Technology, 7(4), 044004.

Salmon, W. C. (1978). Why ask "Why?"? An inquiry concerning scientific explanation. Proceedings and Addresses of the American Philosophical Association, 51(6), 683–705.

Samek, W., & Müller, K. (2019). Towards explainable artificial intelligence. In W. Samek, G. Montavon, A. Vedaldi, L. Hansen, & K. R. Müller (Eds.), Explainable AI: Interpreting, explaining and visualizing deep learning. Lecture Notes in Computer Science (Vol. 11700). Springer.

Sand, M. (2018). The virtues and vices of innovators. In Futures, visions, and responsibility: An ethics of innovation (pp. 241–264).

Sand, M. (2019). On "not having a future". Futures, 107, 98–106. DOI: https://doi.org/10.1016/j.futures.2019.01.002

Sand, M. (2023). The questions of hermeneutic TA: Towards a toolbox. In Hermeneutics, history, and technology: The call of the future (pp. 73–88). Routledge. DOI: https://doi.org/10.4324/9781003322290-5

Sand, M., Grunwald, A., & Nordmann, A. (2023). Hermeneutics, history, and technology: The call of the future. Routledge.

Sayler, K. (2026). Defense primer: Quantum technology (CRS Report No. IF11836). Congressional Research Service. https://www.congress.gov/crs-product/IF11836

Scheffler, S., & Munoz-Dardé, V. (2005). The division of moral labour. Proceedings of the Aristotelian Society, Supplementary Volumes, 229–284.

Schmidt, A., Artaud, A., Aydinoglu, A. U., Bötticher, A., Bravo, R. A., Chiofalo, M., Coates, R., Ercan, I., Grinbaum, A., Haworth, E., Ten Holter, C., de Jong, E., Karstens, B., Hettemann, M., Knörr, A., Ai Ling Lee, C., Marco, F., Mehnert, W., Meyer, J. C., Sharma, S., Vermaas, P., Weidner, C., Wellmann, B., Wolf-Bauwens, M., & Seskir, Z. C. (2025). Current and future directions for responsible quantum technologies: A ResQT community perspective. arXiv preprint arXiv:2509.19815.

Schneider, C., Roßmann, M., Lösch, A., & Grunwald, A. (2021). Transformative vision assessment and 3-D printing futures: A new approach of technology assessment to address grand societal challenges. IEEE Transactions on Engineering Management, 70(3), 1089–1098.

Schutz, G., & Lambert, K. (1994). Outline of a theory of understanding. Synthese, 101, 65–120.

Scott, D. (2011). The technological fix criticisms and the agricultural biotechnology debate. Journal of Agricultural and Environmental Ethics, 24, 207–226.

Selin, C. (2014). On not forgetting futures. Journal of Responsible Innovation, 1(1), 103–108. DOI: https://doi.org/10.1080/23299460.2014.884378

Selin, C., Rawlings, K. C., de Ridder-Vignone, K., Sadowski, J., Altamirano Allende, C., Gano, G., & Guston, D. H. (2017). Experiments in engagement: Designing public engagement with science and technology for capacity building. Public Understanding of Science, 26(6), 634–649.

Seskir, Z. C., Umbrello, S., Coenen, C., & Vermaas, P. E. (2023). Democratization of quantum technologies. Quantum Science and Technology, 8(2), 024005.

Sheikh, H., Prins, C., & Schrijvers, E. (2023). Mission AI: The new system technology. Springer Nature.

Shelley-Egan, C. (2010). The ambivalence of promising technology. NanoEthics, 4(2), 183–189.

Shelley-Egan, C. (2011). Ethics in practice: Responding to an evolving problematic situation of nanotechnology in society [Doctoral thesis]. University of Twente.

Shelley-Egan, C. (2025). The risk of ethicalisation in ethical engagement with quantum technologies: Some brief considerations. NanoEthics, 19(2), 7. DOI: https://doi.org/10.1007/s11569-025-00471-2

Shelley-Egan, C., & De Jong, E. (2025). From nano to quantum: Ethics through a lens of continuity. Science and Engineering Ethics, 31(5), 1–25.

Siffels, L., & Sharon, T. (2024). Where technology leads, the problems follow: Technosolutionism and the Dutch contact tracing app. Philosophy & Technology, 37(4), 125.

Singer, P. (1972). Moral experts. Analysis, 32(4), 115–117.

Singh, H. (2024). Managing the quantum cybersecurity threat: Harvest now, decrypt later. In Quantum computing (pp. 142–158). CRC Press.

Smith III, F. L. (2020). Quantum technology hype and national security. Security Dialogue, 51(5), 499–516.

Sollie, P. (2007). Ethics, technology development and uncertainty: An outline for any future ethics of technology. Journal of Information, Communication and Ethics in Society, 5(4), 293–306.

Sollie, P. (2009). On uncertainty in ethics and technology. In Evaluating new technologies: Methodological problems for the ethical assessment of technology developments (pp. 141–158). Springer Netherlands.

Stahl, B. C., & Flick, C. (2010). ETICA workshop on computer ethics: Exploring normative issues. In Privacy and Identity Management for Life: 6th IFIP WG International Summer School (pp. 64–77). Springer.

Steel, D. (2013). The precautionary principle and the dilemma objection. Ethics, Policy & Environment, 16(3), 321–340.

Steel, D. (2015). Philosophy and the precautionary principle. Cambridge University Press.

Steen, M., Sand, M., & Van de Poel, I. (2021). Virtue ethics for responsible innovation. Business and Professional Ethics Journal.

Stegmaier, P. (2009). The rock 'n' roll of knowledge co-production. Science & Society Series on Convergence Research. EMBO Reports, 10(2), 114.

Stein, S. (2024). Universities confronting climate change: Beyond sustainable development and solutionism. Higher Education, 87(1), 165–183.

Strevens, M. (2013). No understanding without explanation. Studies in History and Philosophy of Science Part A, 44(3), 510–515. DOI: https://doi.org/10.1016/j.shpsa.2012.12.005

Swierstra, T., & Rip, A. (2007). Nano-ethics as NEST-ethics: Patterns of moral argumentation about new and emerging science and technology. NanoEthics, 1(1), 3–20.

Swierstra, T., Stemerding, D., & Boenink, M. (2009). Exploring techno-moral change: The case of the obesity pill. In P. Sollie & M. Düwell (Eds.), Evaluating new technologies: Methodological problems for the ethical assessment of technology developments (pp. 119–138). Springer.

Ten Holter, C., Inglesant, P., Srivastava, R., & Jirotka, M. (2022). Bridging the quantum divides: A chance to repair classic(al) mistakes? Quantum Science and Technology, 7(4), 044006. DOI: https://doi.org/10.1088/2058-9565/ac8db6

Ten Holter, C., Inglesant, P., & Jirotka, M. (2023). Reading the road: Challenges and opportunities on the path to responsible innovation in quantum computing. Technology Analysis & Strategic Management, 35(7), 844–856.

Tepper, A. (1996). Controlling technology by shaping visions. *Policy Sciences*, *29*(1), 29-44.

Tippmann, M., Kaltwasser, J., Mengler, M., et al. (2026). Practical quantum key distribution networks: A mini-review. European Physical Journal Special Topics. DOI: https://doi.org/10.1140/epjs/s11734-026-02316-9

Timmermans, J., & Blok, V. (2021). A critical hermeneutic reflection on the paradigm-level assumptions underlying responsible innovation. *Synthese*, *198*(Suppl 19), 4635-4666. DOI: https://doi.org/10.1007/s11229-018-1839-z

TNO, CWI, & AIVD (2023). The PQC migration handbook: Guidelines for migrating to post-quantum cryptography (Version 2). https://publications.tno.nl/publication/34641918/oicFLj/attema-2023-pqc.pdf

Top, N. (2024). Preparing for a future with responsible quantum technologies: Understanding quantum technologies through a lens of responsible tech.

SURF. https://www.surf.nl/files/2024-10/responsible-quantum-report-surf-preparing-for-a-future-with-responsible-quantum-technologies-october-2024.pdf

Trouwborst, A. (2006). Precautionary rights and duties of states (Vol. 25). Brill.

Umbrello, S. (2024). Ethics of quantum technologies: A scoping review. International Journal of Applied Philosophy.

UNICC (n.d.-a). Quantum for good: Frontier technology for the SDGs [Web page]. https://www.unicc.org/event-quantum-for-good-sep2024/

UNICC (n.d.-b). Quantum for good: Setting the stage for the international year of quantum [Web page]. https://www.unicc.org/quantum-for-good-nov-2024/

Urueña, S. (2022). Responsibility through anticipation? The 'future talk' and the quest for plausibility in the governance of emerging technologies. NanoEthics, 15(3), 271–302.

Urueña, S. (2023). Enacting anticipatory heuristics: A tentative methodological proposal for steering responsible innovation. Journal of Responsible Innovation, 10(1), 2160552.

Urueña, S. (2024). Anticipation and its degrees of critical-reflective radicality: Opening up the affordances of engaging with futures to problematize STI. Journal of Responsible Innovation, 11(1), 2294537.

Vallor, S. (2016). Technology and the virtues: A philosophical guide to a future worth wanting. Oxford University Press.

Valsorda, F. (2026, April 6). A cryptography engineer's perspective on quantum computing timelines. https://words.filippo.io/crqc-timeline/

Van de Merbel, A., Peer, J., Willems, S. J., & Cramer, J. (2024). "Quantum technology will change my life." Citizens' attitudes and knowledge of quantum science and technology. Journal of Physics Communications, 8, 075005.

Van de Poel, I. (2009). Values in engineering design. In A. Meijers (Ed.), Philosophy of technology and engineering sciences (pp. 65–91). Elsevier.

Van de Poel, I. (2016). An ethical framework for evaluating experimental technology. Science and Engineering Ethics, 22(3), 667–686.

Van de Poel, I., & Royakkers, L. (2011). Ethics, technology, and engineering: An introduction. John Wiley & Sons.

Van der Burg, S. (2014). On the hermeneutic need for future anticipation. *Journal of Responsible Innovation, 1*(1), 99–102.

Van der Enden, K. L., Profitiliotis, G., & Croese, D. (2024, September). Advancing a responsible future quantum internet. In 2024 IEEE International Conference on Quantum Computing and Engineering (QCE) (Vol. 2, pp. 65–70). IEEE.

Vaseashta, A. (2025). Dual-use artificial intelligence — Trends, security challenges, and mitigation strategies. In Spectrum of dual-use technologies: Unforeseen risks versus returns (pp. 3–37). Springer Nature Switzerland.

Verbeek, P. P. (2005). What things do: Philosophical reflections on technology, agency, and design. Penn State University Press.

Verbeek, P. P. (2006). Persuasive technology and moral responsibility: Toward an ethical framework for persuasive technologies. Persuasive, 6, 1–15.

Verbeek, P. P. (2010). Accompanying technology: Philosophy of technology after the ethical turn. Techné: Research in Philosophy & Technology, 14(1).

Verhaeghe, P. (2025). Wijsheid. De Bezige Bij.

Vermaas, P. E. (2017). The societal impact of the emerging quantum technologies: A renewed urgency to make quantum theory understandable. Ethics and Information Technology, 19, 241–246. DOI: https://doi.org/10.1007/s10676-017-9429-1

Vermaas, P., & Mans, U. (2024). Quantum technologies and their global impact: Discussion paper. UNESCO HQ. https://unesdoc.unesco.org/ark:/48223/pf0000388955

Vogiatzoglou, P. (2025). The EU's quest for digital sovereignty: A matter of quantum innovation? Digital Society, 4(1), 16.

Von Schomberg, R. (2013). A vision of responsible research and innovation. In R. Owen, J. Bessant, & M. Heintz (Eds.), Responsible innovation: Managing the responsible emergence of science and innovation in society (pp. 51–74). Wiley.

Von Schomberg, R. (2014). The quest for the 'right' impacts of science and technology: A framework for responsible research and innovation. In N. M. de Saille (Ed.), Responsible innovation (pp. 33–50). Wiley.

Weber, E., Reydon, T., Boon, M., & Houkes, W. (2013). The ICE-theory of technical functions. Metascience, 22, 23–44. DOI: https://doi.org/10.1007/s11016-012-9642-9

Weinberg, S. J., Sanches, F., Ide, T., Kamiya, K., & Correll, R. (2023). Supply chain logistics with quantum and classical annealing algorithms. Scientific Reports, 13(1), 4770.

Weslake, B. (2010). Explanatory depth. Philosophy of Science, 77(2), 273–294.

Westerblad, O. (2026). Instrumental understanding. Synthese, 207(1), 44.

Wickson, F., & Carew, A. L. (2014). Quality criteria and indicators for responsible research and innovation: Learning from transdisciplinarity. Journal of Responsible Innovation, 1(3), 254–273.

Wieners, C. E., Hofbauer, B. P., De Vries, I. E., Honegger, M., Visioni, D., Russchenberg, H. W., & Felgenhauer, T. (2023). Solar radiation modification is risky, but so is rejecting it: A call for balanced research. Oxford Open Climate Change, 3(1), kgad002. DOI: https://doi.org/10.1093/oxfclm/kgad002

Wilkenfeld, D. A. (2013). Understanding as representation manipulability. Synthese, 190, 997–1016.

Wilson, A. (2017). Techno-optimism and rational superstition. Techné: Research in Philosophy & Technology, 21, 342–362. DOI: https://doi.org/10.5840/techne201711977

Woodward, J. (2003). Making things happen: A theory of causal explanation. Oxford University Press.

Woodward, J., & Hitchcock, C. (2003). Explanatory generalizations, part I: A counterfactual account. Noûs, 37(1), 1–24.

World Economic Forum (2022). Quantum computing governance principles. WEF. https://www3.weforum.org/docs/WEF_Quantum_Computing_2022.pdf

World Economic Forum (2024a). Quantum for society: Meeting the ambition of the SDGs. https://www.weforum.org/publications/quantum-for-society-fulfilling-the-promise-of-the-sdgs/

World Economic Forum (2024b). Quantum economy blueprint. https://www3.weforum.org/docs/WEF_Quantum_Economy_Blueprint_2024.pdf

WRR (2021). Opgave AI: De nieuwe systeemtechnologie. WRR.

Wynne, B. (2007). Public participation in science and technology: Performing and obscuring a political–conceptual category mistake. *East Asian Science, 1*, 99–110. DOI: https://doi.org/10.1007/s12280-007-9004-7

Yoder, S. D. (1998). The nature of ethical expertise. The Hastings Center Report, 28(6), 11–19.

Zelenovic, J., Taghizadeh, L., Pena-Gonzalez, E., Gómez García, J., & Preneel, B. (2026). Post-quantum discovery as a governance capability: Evidence-based cryptographic visibility and exposure prioritisation in a critical service provider. arXiv. https://arxiv.org/abs/2605.16549

Zhang, X., Zhang, T., Wei, X., Xiao, Z., & Zhang, W. (2024). Reducing potential dual-use risks in synthetic biology laboratory research: A dynamic model of analysis. Humanities and Social Sciences Communications, 11(1), 1–14.

Zhu, Y. (2025). A call for technological understanding. Technology and Language, 6(2), 109–126.

Zwart, H., Landeweerd, L., & Van Rooij, A. (2014). Adapt or perish? Assessing the recent shift in the European research funding arena from 'ELSA' to 'RRI'. Life Sciences, Society and Policy, 10(1), 11.

Summary

Technological Understanding and Ethics Readiness of Technology

Quantum technologies form an emerging group of technologies that harness the principles of quantum mechanics. Motivated by the ambition to do things right from the start, an ethical discussion is emerging around these technologies. Such early-stage or 'upstream' ethics comes with two epistemic challenges: gaining an understanding of a novel technology, and navigating the uncertainties surrounding its development. This thesis develops two concepts that elucidate these challenges and offer tools to navigate them: an account of *technological understanding*, illuminating what it means to understand technology, and the notion of *ethics readiness,* informing which ethical questions are best asked when.

Strictly speaking, a first generation of quantum technologies is already up and running. This includes magnetic resonance imaging (MRI), lasers, and transistors. Now a second generation of quantum technologies is emerging, leveraging actively controlled quantum phenomena for measuring physical data (quantum sensing), performing computational tasks (quantum simulation and computing), and transmitting information (quantum communication). It is this second generation that this thesis focuses on as an illustrative case of ethical engagement with new and emerging technologies. While their stages of development diverge, second-generation quantum technologies share the promise of enabling a wide range of applications across multiple fields – from breakthroughs in the sciences with implications for, for example, the energy transition, to optimising complex logistics and GPS-independent navigation, and from providing novel ways of securing communication to breaking current encryption methods. This promising character, together with the breadth of potential applications, invites ethical engagement: what ethical questions do these technologies raise?

Before turning to that question, this thesis takes a step back to reflect on what enables meaningful engagement with technology in the first place. 'Meaningful' is here understood as ethical discussion grounded in technological reality, focusing on questions that directly arise from what the technology is and may do, to the best of our understanding. Ensuring such groundedness is especially pertinent in the case of new and emerging technologies, given their specific epistemic context: they present us with novel technological capabilities while coming with profound uncertainties surrounding their development. The epistemic conditionality of technology ethics – the idea that ethical engagement with technology is conditioned by epistemic factors – has received relatively little systematic attention. Neglecting the epistemic conditions of

meaningful ethical discussion, however, leaves room for misguided and premature questions that lack grounding in what a technology actually is or may do. To address this gap, this thesis asks:

> *What are the epistemic conditions of meaningful ethical engagement with technology?*

In other words, what do we need to have a grounded discussion about the ethical aspects of a new technology? To investigate this, the thesis zooms in on the twofold epistemic challenge of the ethics of new and emerging technologies: coming to grips with a novel technology, and navigating the deep uncertainties surrounding its development.

The first challenge points towards the need to develop an understanding of the technology. In the absence of a philosophical account of what understanding means in relation to technology, the following sub-question was posed:

> *What does understanding mean in relation to technology, and which form of understanding best enables meaningful ethical analysis?*

The first part of the thesis addresses this question and develops an account of *technological understanding* and its relevance to ethics. **Chapter 2**, *A Conception of Technological Understanding*, lays the theoretical groundwork. Drawing on the notion of scientific understanding, technological understanding is defined as a cognitive skill that is involved whenever we engage with technology. More precisely, it entails the ability to recognise how to use a technological artefact to achieve a practical aim. The account highlights the epistemic dimension of engaging with technology, arguing that any form of successful engagement – whether building it, practically using it, or devising new applications – requires a certain degree and type of understanding. What degree and type depends on the context. To illustrate this, technological understanding is specified for one foundational context: design. Within design, technological understanding amounts to the ability to *design* a technological artefact capable of realising a predefined aim.

Chapter 3, *A Contextual Approach to Technological Understanding*, further develops this contextual character by specifying technological understanding for two further contexts. In the context of operation – in which a technology is practically used – technological understanding entails the ability to recognise how to operate a technological artefact to achieve an aim. In the context of innovation – in which new applications are being devised – it amounts to the ability to recognise a technology's fit for purpose and envision its successful deployment. Together, these three contexts – design, operation, innovation – represent the main forms of engagement with technology. Each context requires the technology to be intelligible in different ways.

The designer needs to grasp the technology's inner workings; the operator its direct consequences; the innovator its functional capabilities. Rather than treating degrees of understanding as varying along a single scale on which some people simply score higher than others, this pluralistic and contextual picture recognises distinct forms of understanding suited to distinct practices, providing a more precise basis for determining which type of understanding is required in a given situation.

Chapter 4, *Technological Understanding and Ethics*, argues that technological understanding is also crucial for ethical engagement with technology. Being ultimately concerned with what a technology does in the world – how it may be used, by whom, for what purposes, and with what consequences – ethics requires grasping a technology's potential uses and applications. This is provided most directly by the type of understanding linked to the innovation context: grasping what the technology is capable of and what it may be used for. Such 'functional' technological understanding offers a direct basis for tracing potential impact and identifying ethical stakes. The claim here is one of relevance rather than exclusivity: functional understanding is essential for meaningful ethical engagement, though other types may also illuminate ethically relevant aspects.

Together, these three chapters answer the central question of the first part of this thesis. They argue that understanding a technology means being able to recognise how to use it to realise an aim. Any form of engaging with technology involves this cognitive skill. Ethical engagement, too, requires it. The contextual character of technological understanding, however, highlights that different contexts require that the technology is intelligible to one at different quantitative and qualitative levels. For ethical purposes specifically, recognising how a technology may be used is essential. It is functional understanding, rather than technical or operational understanding, that most directly provides such insight.

This applies directly to quantum technologies, where introductions for non-experts often focus on the underlying physics rather than the functions. The account of technological understanding developed here provides a basis for redirecting such efforts and, more broadly, for underscoring the importance of understanding quantum technologies as a precondition for meaningful ethical engagement with them.

The second part of the thesis shifts the attention to the other epistemic challenge of the ethics of new and emerging technologies: dealing with the profound uncertainties surrounding their development. At an early-stage of development, much is still unknown – whether the technology will materialise, how it will be used, and with what consequences. Yet early-stage ethical engagement has a clear incentive: a technology that is still taking shape offers many entry points for intervention, making upstream engagement an effective way to steer downstream consequences. This tension is captured by the Collingridge dilemma: early in the innovation process there is still

ample possibility to intervene, yet limited knowledge about what to intervene on; as the technology matures, this knowledge grows but the possibilities for effectively directing its development narrow. While many methods have been developed for ethical analysis of early-stage technologies, an overarching framework that accounts for the fact that the epistemic situation is not static but evolving, remains lacking. It is therefore unclear when which approach to ethics is best suited. This motivates the second sub-question of this thesis:

> *How should we calibrate our ethical engagement to what is – and is not yet – understood about a technology?*

Chapter 5 directly responds to this question. It develops the notion of *ethics readiness* to indicate that a technology allows for different types of ethical scrutiny at different stages of its development. Ethics readiness thus does not reflect a technology's ethical acceptability or the presence of adequate governance structures, but its readiness for specific types of ethical questions. Two main types are distinguished: approaches that focus on the attribution of meaning to a technology – scrutinising how it is envisioned and by whom – and approaches that focus directly on its implications, assessing potential outcomes. These approaches deal with uncertainty differently: the meaning-oriented approach shifts focus from the future technology to present ideas about it, while the outcomes-oriented approach directly engages with potential futures and their implications.

The chapter argues that the appropriateness of these methods shifts as a technology matures and its epistemic context changes. In other words, the ethics readiness of a technology evolves along with technological development. Drawing on this idea, a framework is proposed that guides which ethical questions are best asked when. The meaning-oriented approach is best suited to deal with the profound uncertainties at the earliest stages of development, whereas the outcomes-oriented approach risks speculative overreach. As the technology further develops and takes shape more concretely, the outcomes-oriented approach gains both feasibility and relevance. The ethics readiness framework thus links the appropriateness of ethical approaches to a technology's epistemic context.

Applied to quantum technologies, the framework suggests a different approach for quantum sensing and for quantum computing and communication, accounting for their different of development: quantum sensing being already operational, while computing and communication remain largely conceptual and early experimental. While quantum sensing may thus be ready for outcomes-oriented ethics, quantum computing and communication are best approached through meaning-oriented ethics.

Chapter 6 takes up this implication. It adopts a meaning-oriented approach to the ethical analysis of quantum computing. It does so by examining the visions of key

actors in the quantum computing ecosystem, asking how they envision the future of this technology. Based on an empirical analysis of their communications about what quantum computing will be and do, four main visions emerge. Each of them frames quantum computing as a powerful tool that will bring about change, yet situate this change in different domains: as an effective contributor to societal benefit, as a groundbreaking scientific tool, as an important economic asset, and as a strategic geopolitical advantage.

Scrutinising these visions from an ethical perspective reveals several issues warranting attention: a pervasive techno-optimism paired with technosolutionist reasoning; promissory language used to legitimise innovation agendas; and a discourse shaped by the interests of a relatively narrow group of powerful actors. If anything, these findings highlight the risk of a discourse dominated by a narrow elite, and the importance of opening it up to broader societal scrutiny and debate about the values and priorities guiding quantum development.

Chapter 7 takes up one of the flagged issues: technosolutionism. It identifies the phenomenon of *Quantum for Good* – the aspiration to mobilise quantum technologies for societal benefit – as exhibiting a technosolutionist logic: presenting technology as an effective solution to social problems. So understood, technosolutionism is not inherently problematic. Yet it does carry ethical risks, primarily by backgrounding alternatives and obscuring morally relevant aspects of the problem a technosolution seeks to address. This chapter argues that assessing the ethical acceptability of technosolutionist approaches requires scrutinising the construction of the problem, its sociopolitical dimensions, the impact of the proposed solution, and the plausibility of its claimed effectiveness.

Assessed against these criteria, the technology-first orientation of Quantum for Good carries a real risk: social challenges may be reframed in narrow, technically tractable terms, depoliticising inherently political questions and marginalising more effective alternatives. While Quantum for Good is laudable in its commitment to socially engaged innovation, it must attend to the limits of technological fixes and acknowledge the social and political dimensions of the challenges it seeks to address – if it aspires to be more than a moral veneer for existing innovation agendas.

Chapter 8, examines a 'quantum for bad' scenario: the potential of a full-scale quantum computer to break widely used encryption methods, posing an unprecedented cyber security threat. This scenario embodies the most tangible and potentially urgent ethical implication of quantum computing. However, a full-scale quantum computer capable of materialising this threat does not yet exist and its feasibility is not uncontested. This situation challenges the ethics readiness framework developed in Chapter 5, which would suggest that an outcomes-oriented approach is premature at this stage of quantum computing's development. Yet waiting for the technology to mature

sufficiently may foreclose effective anticipatory interventions. The chapter therefore asks under what conditions engaging with potential outcomes can be justified, even when the ethics readiness framework would deem this premature.

The chapter argues that outcomes-oriented ethical engagement is justified when the claimed risk can be made sufficiently plausible – that is, when there is adequate grounding for treating a speculative development as a genuine threat warranting ethical attention. Establishing this plausibility then raises a further question: when is precautionary action ethically defensible in response to such a risk? The chapter argues that this requires not only a plausible risk claim but also a compelling precautionary response; one that can be shown to be both effective and proportionate given the costs and uncertainties involved. The quantum cyber security case satisfies both conditions. The analysis provides the dominant narrative with a more ethically robust grounding, and by making its underlying assumptions explicit, it opens them up to broader contestation and scrutiny.

Together, the chapters of the second part show how ethical engagement with early-stage technologies can take shape while taking seriously the uncertainties that surround their development. The applied chapters also directly contribute to the emerging discussion on the ethical aspects of quantum technologies, and quantum computing in particular.

While prompted by the emergence of quantum technologies, this thesis engages with challenges that any ethical engagement with new and emerging technologies faces: coming to grips with a novel technology and navigating the profound uncertainties surrounding its development. Meaningful ethical engagement hinges on addressing both, focusing on issues that gain their relevance and urgency from a grounded understanding of what the technology is and can plausibly be expected to do. The conceptual tools of technological understanding and ethics readiness help guide such engagement and illuminate what it means to ask the right questions at the right moment.

Samenvatting

Technologisch Begrip en Ethische Ontvankelijkheid van Technologie

Quantumtechnologieën vormen een familie van opkomende technologie gestoeld op quantum mechanica. Gedreven door de ambitie om het dit keer van meet af aan goed aan te pakken is er in een vroeg stadium een ethische discussie over deze technologieën op gang gekomen. Zulke vroegtijdige of 'stroomopwaartse' ethiek van technologie – waarbij ethische vragen al bij de bron van technologische ontwikkeling worden gesteld – kent twee uitdagingen die te maken hebben met kennis. Allereerst moet er begrip worden ontwikkeld van de nieuwe technologie: we moeten begrijpen wat de technologie inhoudt om de ethische aspecten ervan te kunnen inzien. Daarnaast moeten we zien om te gaan met het gebrek aan kennis rondom een nieuwe technologie. De ontwikkeling van technologie kent allerlei onzekerheden waar we rekenschap van moeten geven in ethische discussies. In dit proefschrift worden twee concepten ontwikkeld die helpen om grip te krijgen op deze uitdagingen. Het eerste is *technologisch begrip*, het soort begrip dat we nodig hebben in onze omgang met technologie. Het tweede is *ethische ontvankelijkheid*, het idee dat de ontvankelijkheid van technologie voor bepaalde typen ethische vragen verandert naarmate de technologie volwassener wordt. Door deze twee begrippen toe te passen op quantumtechnologieën laat dit proefschrift zien hoe we over een technologie die nog in de kinderschoenen staat toch zinnige ethische vragen kunnen stellen.

Een eerste generatie quantumtechnologieën heeft al zijn weg naar de samenleving gevonden. Hieronder vallen magnetic resonance imaging (MRI), lasers, en transistors. Nu maakt een tweede generatie zijn opwachting, waarbij fenomenen uit de quantum mechanica dusdanig worden gecontroleerd dat ze kunnen worden gebruikt om fysieke data te meten ('quantum sensing'), om berekeningen uit te voeren ('quantum computing'), en om informatie over te brengen ('quantum communicatie'). Het is deze tweede generatie quantumtechnologieën die in dit proefschrift model staat voor ethische discussies rondom opkomende technologie. Hoewel de nieuwe generatie quantumtechnologieën zich in verschillende ontwikkelingsfasen bevinden, delen ze de belofte een heel scala aan toepassingen mogelijk te maken – van doorbraken in de wetenschap met implicaties voor de energietransitie, tot het optimaliseren van complexe logistiek en navigeren zonder GPS, en van nieuwe manieren om communicatie te beveiligen tot het breken van huidige encryptie. Dit veelbelovende karakter en de breedte aan potentiële toepassingen nodigt uit tot reflectie: welke ethische vragen werpen deze technologieën op?

Voordat we met die vraag aan de slag kunnen is het van belang om stil te staan bij de vraag wat er nodig is voor een zinvolle ethische discussie. 'Zinvol' in de zin dat de discussie gegrond is in de realiteit van de technologie, met vragen die voortkomen uit wat de technologie is en kan. Erop toezien dat een ethische discussie gegrond is, is bij uitstek belangrijk bij nieuwe technologieën, gegeven hun specifieke context: ze brengen allerlei nieuwe mogelijkheden met zich mee en zijn daarnaast omgeven met onzekerheden over hun ontwikkeling en consequenties. Deze epistemische (kennisgerelateerde) dimensie van de ethiek rondom nieuwe technologie is relatief onderbelicht. Als deze aspecten van 'nieuwigheid' en 'onzekerheid' echter worden genegeerd, vervaagt het onderscheid tussen gefundeerde ethische vragen en vragen die ongegrond of prematuur zijn. In dit proefschrift staat de volgende vraag daarom centraal:

> *Wat zijn de epistemische voorwaarden van een betekenisvolle ethische discussie over technologie?*

Met andere woorden, wat is er nodig voor een gegrond gesprek over de ethische aspecten van nieuwe technologie? Om dit te onderzoeken richt dit proefschrift zich op de tweevoudige epistemische uitdaging van de ethiek van nieuwe technologie: een nieuwe technologie leren kennen en omgaan met de vele onzekerheden rondom de ontwikkeling ervan.

De eerste uitdaging ligt in het ontwikkelen van een begrip van de technologie. Omdat een duidelijk idee van wat het betekent om een technologie te begrijpen ontbreekt in de techniekfilosofie, wordt de volgende subvraag gesteld:

> *Wat betekent het om een technologie te begrijpen en welke vorm van begrip maakt een gegronde ethische discussie mogelijk?*

Het eerste deel van dit proefschrift ontwikkelt het idee van *technologisch begrip* en de relevantie ervan voor ethiek. **Hoofdstuk 2**, *Een Definitie van Technologisch Begrip*, legt hiervoor het theoretische fundament. Voortbordurend op het idee van wetenschappelijk begrip, wordt technologisch begrip hier gedefinieerd als een cognitieve vaardigheid die je nodig hebt in de omgang met technologie. Preciezer gezegd is het de vaardigheid om een technologie te kunnen gebruiken om daarmee een doel te behalen. Het idee van technologisch begrip laat zien dat onze omgang met technologie een cognitieve dimensie heeft: elk succesvol gebruik – of dat nu gaat om het bouwen ervan, het praktisch gebruiken, of het bedenken van nieuwe toepassingen – vergt een zekere mate van begrip. Hoeveel en wat voor begrip precies hangt af van de context. Om dit te illustreren wordt het idee van technologisch begrip verder uitgewerkt voor een specifieke context: design. Binnen deze context komt technologisch begrip neer op de vaardigheid om een technologie te ontwerpen die in staat is om een vooraf bepaald doel te realiseren.

Hoofdstuk 3, *Een Contextuele Benadering van Technologisch Begrip*, ontwikkelt dit contextuele karakter verder en specificeert het begrip voor twee aanvullende contexten. In de context van 'operatie' – waarin een technologie praktisch wordt aangewend – behelst technologisch begrip de vaardigheid om in te zien hoe de technologie moet worden gebruikt om een doel te bereiken. In de context van 'innovatie' – waarin nieuwe toepassingen worden bedacht – houdt technologisch begrip de vaardigheid in om in te schatten voor welke doelen een technologie succesvol kan worden ingezet. Samen vormen deze drie contexten – design, operatie en innovatie – de drie hoofdvormen van omgang met technologie. Elke context vergt een ander type begrip, waarbij de technologie op een andere manier inzichtelijk is voor de gebruiker. De designer dient de interne werking van de technologie te begrijpen; de praktische gebruiker moet inzicht hebben in de directe gevolgen ervan; en de innovator moet de functionele capaciteiten ervan inzien. Technologisch begrip laat zich dus niet alleen uitdrukken in gradaties, maar kent ook verschillende vormen. Deze pluralistische kijk maakt het mogelijk om verschillende vormen van technologisch begrip te koppelen aan verschillende contexten, en laat ruimte voor verschillende soorten expertise. Het geeft daarmee een preciezer uitgangspunt voor het ontwikkelen en testen van begrip van technologie.

Hoofdstuk 4, *Technologisch Begrip en Ethiek*, laat zien dat ethische discussies ook voortbouwen op technologisch begrip. Ethiek houdt zich bezig met wat een technologie doet in de wereld: hoe technologie wordt gebruikt, door wie, waarvoor, en met welke gevolgen. Daarvoor is inzicht nodig in het mogelijke gebruik en toepassingen van een technologie. Dit inzicht wordt het meest direct verschaft door het type begrip dat van toepassing is in een innovatiecontext: herkennen waartoe de technologie in staat is en waarvoor ze ingezet kan worden. Dit soort 'functioneel' begrip (gericht op mogelijke functies) biedt een directe basis voor het inschatten van potentiële impact en het identificeren van ethische aspecten. Deze claim gaat meer over relevantie dan exclusiviteit: functioneel begrip van technologie speelt een sleutelrol bij een gegronde ethische discussie, maar dat sluit niet uit dat andere vormen van technologisch begrip ook licht kunnen werpen op andere ethische aspecten.

Samen beantwoorden deze drie hoofdstukken de onderzoeksvraag van het eerste deel van dit proefschrift. Het begrijpen van een technologie betekent in staat zijn om te herkennen hoe het kan worden gebruikt om een doel te bereiken. Deze cognitieve vaardigheid speelt een rol in elke omgang met technologie – ook in ethische discussies over technologie. Het contextuele karakter van technologisch begrip benadrukt echter dat verschillende contexten vereisen dat de technologie op kwantitatief en kwalitatief verschillende niveaus begrijpelijk is. Voor ethische doeleinden specifiek is het cruciaal te herkennen hoe een technologie gebruikt kan worden: een functiegericht begrip, eerder dan een technisch of operationeel begrip, is het meest direct relevant. Dit inzicht is direct van toepassing op quantumtechnologieën, waarbij introducties voor niet-experts zich vaak richten op de onderliggende fysica in plaats van op de functies. De hier

ontwikkelde theorie van technologisch begrip biedt een basis om dergelijke inspanningen bij te sturen en, meer in het algemeen, om het belang te onderstrepen van het begrijpen van quantumtechnologieën als voorwaarde voor zinvolle ethische omgang ermee.

Het tweede deel van dit proefschrift richt de aandacht op de andere epistemische uitdaging van ethiek van nieuwe technologieën: de vele onzekerheden rondom hun ontwikkeling. In een vroeg stadium is er nog veel onbekend. Het kan onzeker zijn of een technologie er daadwerkelijk komt, in welke vorm, hoe de technologie gebruikt zal worden en wat de gevolgen daarvan zullen zijn. Tegenover die onzekerheden staat dat vroegtijdige ethische discussie een groot voordeel heeft: een technologie die nog in ontwikkeling is biedt ruimte voor effectieve interventies: door 'stroomopwaarts' in te grijpen kunnen de gevolgen 'stroomafwaarts' worden beïnvloed.

Dit spanningsveld wordt gevat in het dilemma van Collingridge: vroeg in het innovatieproces is er nog volop mogelijkheid om in te grijpen, maar is er beperkte kennis over waarop in te grijpen; naarmate de technologie volwassener wordt neemt deze kennis toe maar worden tegelijkertijd de mogelijkheden om haar ontwikkeling effectief te sturen kleiner. Er zijn meerdere methodes ontwikkeld voor de ethische studie van technologieën die zich nog in een vroege ontwikkelingsfase bevinden. Wat nog ontbreekt is een overkoepelend kader dat sturing geeft aan wanneer welke methode het meest geschikt is, gegeven het feit dat technologieën in ontwikkeling zijn en dus ook de kennis erover verandert. Om hier inzicht in te krijgen richt het tweede deel zich op de volgende subvraag:

> *Hoe kunnen we ethische discussie afstemmen op wat we wel – en nog niet – begrijpen van een technologie?*

Hoofdstuk 5 gaat direct in op deze vraag. In dit hoofdstuk wordt het begrip *ethische ontvankelijkheid* gemunt om aan te geven dat een technologie in verschillende ontwikkelingsfasen ontvankelijk is voor verschillende typen ethisch onderzoek. Ethische ontvankelijkheid duidt dus aan in hoeverre een technologie 'klaar' is voor bepaalde ethische vragen. Dit hoofdstuk maakt onderscheid tussen twee typen ethisch onderzoek: benaderingen die zijn gericht op de betekenis die wordt toegekend aan een technologie en benaderingen die zich richten op de implicaties van een technologie. Het eerste type – de betekenisbenadering – houdt zich bezig met hoe een technologie wordt voorgesteld en door wie; het tweede type – de uitkomstenbenadering – kijkt naar mogelijke gevolgen. Deze benaderingen gaan op verschillende manieren om met het probleem van onzekerheid: de betekenisbenadering richt de aandacht op huidige ideeën over een toekomstige technologie, waar de uitkomstenbenadering zich richt op mogelijke toekomsten.

In dit hoofdstuk wordt beargumenteerd dat de geschiktheid van deze benaderingen verschuift naarmate een technologie volwassener wordt en de epistemische context – wat er wel en niet bekend over is – verandert. Met andere woorden, de ethische ontvankelijkheid van een technologie evolueert samen met de technologie zelf. Dit idee leidt tot een denkkader om na te gaan wanneer welke ethische vragen het best kunnen worden gesteld. De betekenisbenadering is het best uitgerust om om te gaan met de vele onzekerheden in de vroegste ontwikkelingsfasen, waar de uitkomstenbenadering al snel het risico loopt te speculatief te worden. Naarmate de technologie zich verder ontwikkelt en concretere vormen aanneemt wint de uitkomstenbenadering aan uitvoerbaarheid en relevantie. Het idee van ethische ontvankelijkheid verbindt de geschiktheid van ethische benaderingen dus direct aan de epistemische context van een technologie.

Toegepast op quantumtechnologieën suggereert het denkkader een andere benadering voor quantum sensing enerzijds en quantum computing en communicatie anderzijds. Deze technologieën bevinden zich immers in verschillende ontwikkelingsstadia: waar quantumsensoren al operationeel zijn, bevinden quantumcomputers en quantumcommunicatie zich nog in een vroege experimentele fase. De eerste is dus al ontvankelijk voor uitkomsten-georiënteerde ethiek, terwijl de laatste twee beter benaderd kunnen worden vanuit een betekenis-georiënteerde ethiek.

Hoofdstuk 6 gaat aan de slag met deze implicatie en kiest een betekenis-georienteerde benadering van ethiek bij de bestudering van quantumcomputers. Het hoofdstuk onderzoekt hoe sleutelactoren binnen het quantum computing ecosysteem zich de toekomst van deze technologie voorstellen. Op basis van een empirische analyse van hun uitlatingen over de toekomst van quantum computing worden vier visies geïdentificeerd. In elke visie wordt quantum computing voorgesteld als een krachtig middel dat verandering teweeg zal brengen binnen een specifiek domein in de samenleving: als een effectief middel om maatschappelijk welzijn te verbeteren, als een baanbrekend wetenschappelijk instrument, als een belangrijke economische troef, en als een strategisch geopolitiek voordeel.

Een ethische analyse van deze visies belicht een aantal aspecten dat verdere aandacht verdient, te weten: een alomtegenwoordig techno-optimisme dat hand-in-hand gaat met een technosolutionistische tendens; beloftevolle taal die wordt gebruikt om innovatieagenda's te legitimeren; en een discours dat wordt gevormd door de belangen van een relatieve kleine groep machtige actoren. Wat deze bevindingen in ieder geval laten zien is dat er een risico heerst van een discours dat wordt gedomineerd door een kleine elite, en het belang om dit te openen voor een breder maatschappelijk debat over de waarden en prioriteiten die de ontwikkeling van quantumtechnologieën sturen.

Hoofdstuk 7 gaat in op één van de gesignaleerde kwesties uit het vorige hoofdstuk: technosolutionisme. Het fenomeen van *Quantum for Good* – de ambitie om quantumtechnologieën in te zetten voor maatschappelijk belang – volgt een technosolutionistische logica: technologie wordt gepresenteerd als een effectieve oplossing voor sociale problemen. Zo begrepen is technosolutionisme niet inherent problematisch. Toch brengt het ethische risico's met zich mee, hoofdzakelijk doordat het alternatieven naar de achtergrond dringt en moreel relevante aspecten van het probleem dat een techno-oplossing wil oplossen, versluiert. Om de ethische aanvaardbaarheid van technosolutionistische benaderingen te beoordelen, is het cruciaal de constructie van het probleem, de sociopolitieke dimensies ervan, de impact van de voorgestelde oplossing en de plausibiliteit van de geclaimde effectiviteit te onderzoeken.

Afgezet tegen deze criteria draagt de technologiegestuurde oriëntatie van *Quantum for Good* – waarbij de technologie als startpunt wordt genomen voor het identificeren en oplossen van problemen – een reëel risico: sociale uitdagingen kunnen worden vertaald naar nauw geformuleerde, technisch hanteerbare problemen, waarmee normatieve vraagstukken worden gedepolitiseerd en buiten het domein van de ethiek worden geplaatst, en effectievere alternatieve naar de achtergrond kunnen verdwijnen. Hoewel het idee achter *Quantum for Good* – het stimuleren van innovatie met een positieve maatschappelijke bijdrage – te prijzen valt, is het cruciaal om de grenzen van technologische oplossingen te erkennen en de sociale en politieke dimensies van de uitdagingen die het poogt te adresseren onder ogen te komen – tenminste, als het méér wil zijn dan een moreel vernislaagje op innovatieagenda's.

Hoofdstuk 8, tot slot, onderzoekt een '*Quantum for Bad*' scenario: het gebruik van een quantum computer om veelgebruikte vormen van encryptie te kraken. Dit scenario zou een ongekende cyberveiligheidsdreiging betekenen, en vormt een van de meest tastbare en potentieel urgente ethische implicaties van quantum computing. Een grootschalige quantumcomputer die deze dreiging kan verwerkelijken bestaat echter nog niet, en de haalbaarheid ervan is niet onomstreden. Deze situatie – van een nog speculatieve technologie en het nadenken over concrete ethische gevolgen – daagt het eerdere denkkader van ethische ontvankelijkheid uit. Immers, volgens dat kader zou uitkomsten-georiënteerde ethiek in het geval van quantum computing voorbarig zijn. Toch kan afwachten in dit geval desastreus zijn omdat het effectieve interventies onmogelijk maakt. Het hoofdstuk vraagt daarom onder welke voorwaarden omgang met mogelijke uitkomsten gerechtvaardigd kan worden, ook wanneer het kader van ethische gereedheid dit in eerste instantie als voorbarig zou aanmerken.

Het hoofdstuk betoogt dat uitkomstengerichte ethiek gerechtvaardigd is wanneer het geclaimde risico aannemelijk genoeg gemaakt kan worden. Dat wil zeggen, wanneer er voldoende grondslag is om een speculatieve ontwikkeling te behandelen als een echte

dreiging die ethische aandacht verdient. Het vaststellen van deze aannemelijkheid roept vervolgens een verdere vraag op: wanneer is voorzorgsactie ethisch verdedigbaar als reactie op een dergelijk risico? Naast een aannemelijk risico vereist dit een overtuigend voorstel voor een voorzorgsmaatregel: één waarvan zowel de effectiviteit als de proportionaliteit aannemelijk kan worden gemaakt, gegeven de kosten ervan en de onzekerheden die meespelen. De dreiging van quantum computing voor cyberveiligheid voldoet aan beide voorwaarden. Daarmee geeft dit hoofdstuk het huidige narratief een ethisch robuustere grondslag. Door de onderliggende aannames daarbij expliciet te maken, opent de analyse bovendien ruimte voor bredere toetsing en discussie.

De hoofdstukken uit dit tweede deel laten zien hoe ethische reflectie op technologieën in een vroeg stadium plaats kan vinden zonder daarbij de onzekerheden rondom technologische ontwikkeling uit het oog te verliezen. De toegepaste hoofdstukken dragen daarnaast direct bij aan de opkomende discussie over de ethische aspecten van quantumtechnologieën en van quantumcomputers in het bijzonder.

Hoewel quantumtechnologieën de aanleiding vormen voor dit proefschrift zijn de uitdagingen die hierin centraal staan van toepassing op alle nieuwe en opkomende technologieën: het leren kennen van een nieuwe technologie om er de juiste vragen over te kunnen stellen, en het omgaan met alle onzekerheden rondom de ontwikkeling ervan zonder passief te worden. Deze twee uitdagingen het hoofd bieden is cruciaal voor betekenisvolle ethische discussies over vragen die voortkomen uit een gegrond begrip van wat de technologie is en kan. De ideeën van *technologisch begrip* en *ethische ontvankelijkheid* van technologie geven handvatten voor deze discussies en helpen om te bepalen wat de juiste vragen zijn op welk moment.

Abstract

Quantum technologies form a new and rapidly developing family of technologies with wide-ranging potential applications – from breakthroughs in materials science and drug discovery to GPS-independent navigation and novel ways of securing or breaking digital communications. Their emergence has prompted an ethical discussion motivated by the ambition to do things right from the start. Such early-stage engagement with the ethical dimensions of technology, however, comes with a distinctive epistemic challenge: how can we ask meaningful ethical questions about a technology we are only beginning to understand, under conditions of profound uncertainty about what it will ultimately be capable of?

This thesis investigates the epistemic conditions of meaningful ethical engagement with technology, using quantum technologies as a paradigm case of new and emerging science and technology. It argues that meaningful ethical discussion – discussion grounded in what a technology actually is and may do – requires two conditions: sufficient understanding of the technology and appropriate timing of ethical engagement. Two concepts are developed to elucidate these conditions: *technological understanding* and *ethics readiness*.

The first part develops a philosophical account of technological understanding. Drawing on the philosophy of science, technological understanding is defined as a cognitive skill: the ability to recognise how to use a technological artefact to achieve an aim. The account is contextual: different forms of engagement with technology (design, operation, innovation) require different types and degrees of understanding. For ethical engagement specifically, it is understanding of a technology's functional capabilities – what it can do and may be used for – that is most directly relevant. This functional understanding provides the basis for tracing potential impact and identifying ethical stakes.

The second part develops the framework of ethics readiness, which holds that the appropriateness of different ethical approaches depends on the maturity of the technology in question. Meaning-oriented approaches – examining the visions and expectations attributed to a technology – are best suited to early-stage development, while outcomes-oriented approaches gain feasibility and relevance as the technology matures. This framework is applied to quantum computing, quantum sensing, and quantum communication, and put to work in ethical analyses of visions of quantum computing, the Quantum for Good initiative, and the quantum cyber security threat.

Together, these two concepts illuminate the epistemic conditionality of ethics of technology and provide the conceptual tools for asking the right ethical questions at the right moment – keeping ethical engagement grounded in technological reality while navigating the uncertainties that inevitably surround new and emerging technologies.

Appendix

Appendix to Chapter 6. Overview of selected actors and materials ('vision statements')

	Actor	Actor type	Vision document	Document type
(1)	Australia	Government	National Quantum Strategy	National strategy
(2)	Austria	Government	Quantum Austria	Research agenda
(3)	Brazil	Government	Competence Center in Quantum Technologies	National action plan
(4)	Canada	Government	National Quantum Strategy	National strategy
(5)	China	Government	China's 14th Five-Year Plan	National strategy
(6)	Denmark	Government	National Strategy for Quantum Technology	National strategy
(7)	European Union	Government	Digital strategy	Supranational strategy
(8)	Finland	Government	Quantum Technology Strategy	National strategy
(9)	France	Government	National Quantum Strategy	National strategy
(10)	Germany	Government	Quantum Technologies Conceptual Framework Programme	National action plan
(11)	Germany	Government	Roadmap for quantum computing	National road map
(12)	Switzerland	Government	Swiss Quantum Initiative	National initiative
(13)	Israel	Government	Israel Nat. Quantum Initiative	National initiative
(14)	India	Government	National Quantum Mission	National strategy
(15)	Netherlands	Government	National Agenda on Quantum Technology	National strategy
(16)	United States	Government	National Quantum Initiative	National strategy
(17)	United Kingdom	Government	National Quantum Strategy	National strategy
(18)	Sweden	Government	Swedish Quantum Agenda	National strategy
(19)	Japan	Government	Strategy of Quantum Future Industry Development	National strategy
(20)	South Korea	Government	Nat. Strategic Plan for Quantum Science and Techn.	National strategy
(21)	Italy	Government	Quantum Technologies: Italian Strategy Document	National strategy
(22)	Ireland	Government	Quantum 2030-A national quantum strategy for Ireland	National strategy
(23)	Lithuania	Government	Lithuanian Quantum Agenda	National agenda
(24)	Spain	Government	Quantum Spain	National strategy
(25)	IBM Quantum (IBM)	Industry	ibm.com	Roadmap
(26)	Google Quantum AI (Alphabet)	Industry	quantumai.google	Website
(27)	Microsoft Azure	Industry	quantum.microsoft.cin	Website / roadmap
(28)	Amazon Braket (AWS)	Industry	aws.amazon.com	Website
(29)	Intel Labs (Intel)	Industry	intel.com	Website

(30)	D-Wave	Industry	dwavequantum.com	Website / e-book
(31)	Rigetti	Industry	rigetti.com	Website
(32)	IonQ	Industry	ionq.com	Website
(33)	Quantinuum	Industry	quantinuum.com	Website
(34)	Infleqtion	Industry	infleqtion.com	Website
(35)	G7	Coordinator	g7.canada.ca	Website
(36)	Open Quantum Institute	Coordinator	open-quantum-institute.cern	Website
(37)	OECD	Coordinator	oecd.org	Paper
(38)	Quantum Delta NL	Coordinator	quantumdelta.nl	Website
(39)	United Nations	Coordinator	quantum2025.org	Website
(40)	National Quantum Computing Centre UK	Coordinator	nqcc.ac.uk	Website
(41)	World Economic Forum	Coordinator	weforum.org	Report
(42)	NATO	Coordinator	nato.int	Website
(43)	Munich Quantum Valley	Coordinator	munich-quantum-valley.de	Website
(44)	EU Quantum Flagship	Coordinator	qt.eu	Website
(45)	QuTech (TU Delft)	Research	qutech.nl	Website
(46)	IQC (University of Waterloo)	Research	uwaterloo.ca	Website
(47)	Harvard Quantum Initiative (Harvard University)	Research	quantum.harvard.edu	Website
(48)	ETH Zurich – Quantum Center	Research	qc.ethz.ch	Website
(49)	Oxford Quantum Institute (Oxford University)	Research	oqi.ox.ac.uk	Website
(50)	JQI (University of Maryland & NIST)	Research	jqi.umd.edu	Website
(51)	Hefei Institutes for Physical Science (Chines Academy of Sciences, China)	Research	hf.cas.cn	Website
(52)	Quantum Physics and Quantum Information dep. (University of Science and Technology, China)	Research	quantum.ustc.edu.cn	Website
(53)	Centre for Quantum Technologies (National University of Singapore)	Research	cqt.sg	Website
(54)	RIKEN Center for Quantum Computing (Japan)	Research	riken.jp	Website
(55)	Institut Quantique (Canada)	Research	usherbrooke.ca	Website
(56)	QuSoft (University of Amsterdam, The Netherlands)	Research	qusoft.org	Website
(57)	Simons Institute (Berkeley University, USA)	Research	simons.berkeley.edu	Website
(58)	IQIM (CalTech, USA)	Research	iqim.caltech.edu	Website
(59)	Scott Aaronson (Quantum Inform. Center, University of Texas)	Thought leader	scottaaronson.blog	Blog
(60)	John Preskill (California Institute of Technology)	Thought leader	scienceexchange.caltech.edu	Interview
(61)	Michele Mosca (Inst. for Quantum Computing, University of Waterloo)	Thought leader	ieeexplore.ieee.org	Paper
(62)	Peter Shor (Massachusetts Institute of Technology)	Thought leader	ep-news.cern.ch	Interview
(63)	Aram Harrow (MIT's Center for Theoretical Physics)	Thought leader	medium.com	Interview
(64)	Artur Eckert (Mathematical Institute, University of Oxford)	Thought leader	executivemagazine.pl	Interview

Other publications by the author

De Jong, E. (2026). From Blind Spot to Stakeholder: Animals in Technology Ethics and Responsible Innovation. *Science & Engineering Ethics*. DOI: https://doi.org/10.1007/s11948-026-00617-9

De Jong, E., & Bossert, L. (2026). Animal and Technology Ethics: The Need to Bring Them into Conversation. Ethics, Policy & Environment, p. 1-6. DOI: https://doi.org/10.1080/21550085.2026.2640852

Lane, B., Boshuijzen-Van Burken, C., **de Jong, E.**, Mittal, A., Sand, M., Seskir, Z., Shelley-Egan, C., & Vermaas, P. (2026). Ethical innovation of quantum computing in Europe: a policy intervention. Journal *of Responsible Technology*, 100188. DOI: https://doi.org/10.1016/j.jrt.2026.100188

Schmidt, A., Artaud, A., Aydinoglu, A. U., Bötticher, A., Araiza Bravo, R., Chiofalo, M., Coates, R., Ercan, İ., Grinbaum, A., Haworth, E., Ten Holter, C., **de Jong, E.**,. . . Seskir, Z. (2026). Current and future directions for responsible quantum technologies: A ResQT community perspective. *Ethics & Society, 20*(3), 28. https://doi.org/10.1007/s11569-026-00522-2

Shelley-Egan, C., & **de Jong, E**. (2025). From nano to quantum: Ethics through a lens of continuity. *Science and Engineering Ethics, 31*(5), p. 31. DOI: https://doi.org/10.1007/s11948-025-00557-w

Gasser, U., **de Jong, E.**, & Kop, M. (2024). A call for responsible quantum technology. *Nature Physics, 20*(4), p. 525-527. DOI: https://doi.org/10.1038/s41567-024-02462-8

Kop, M., Aboy, M., **de Jong, E.**, Gasser, U., Minssen, T., Cohen, I. G., Brongersma, M., Quintel, T., Floridi, L., & Laflamme, R. (2024). Ten principles for responsible quantum innovation. *Quantum Science and Technology, 9*(3), p. 035013. DOI: https://doi.org/10.1088/2058-9565/ad3776

Kop, M., Aboy, M., **de Jong, E.**, Gasser, U., Minssen, T., Cohen, I. G., Brongersma, M., Quintel, T., Floridi, L., & Laflamme, R. (2023). Towards responsible quantum technology: Safeguarding, engaging and advancing quantum R&D. *UC Law Science and Technology Journal, 15*(1), p. 63.

De Jong, E. (2022). Own the unknown: An anticipatory approach to prepare society for the quantum age. *Digital Society, 1*(2), p. 15. DOI: https://doi.org/10.1007/s44206-022-00020-4

Titles in the ILLC Dissertation Series

ILLC DS-2021-11: **Sirin Botan**
Strategyproof Social Choice for Restricted Domains

ILLC DS-2021-12: **Michael Cohen**
Dynamic Introspection

ILLC DS-2021-13: **Dazhu Li**
Formal Threads in the Social Fabric: Studies in the Logical Dynamics of Multi-Agent Interaction

ILLC DS-2021-14: **Álvaro Piedrafita**
On Span Programs and Quantum Algorithms

ILLC DS-2022-01: **Anna Bellomo**
Sums, Numbers and Infinity: Collections in Bolzano's Mathematics and Philosophy

ILLC DS-2022-02: **Jan Czajkowski**
Post-Quantum Security of Hash Functions

ILLC DS-2022-03: **Sonia Ramotowska**
Quantifying quantifier representations: Experimental studies, computational modeling, and individual differences

ILLC DS-2022-04: **Ruben Brokkelkamp**
How Close Does It Get?: From Near-Optimal Network Algorithms to Suboptimal Equilibrium Outcomes

ILLC DS-2022-05: **Lwenn Bussière-Carae**
No means No! Speech Acts in Conflict

ILLC DS-2022-06: **Emma Mojet**
Observing Disciplines: Data Practices In and Between Disciplines in the 19th and Early 20th Centuries

ILLC DS-2022-07: **Freek Gerrit Witteveen**
Quantum information theory and many-body physics

ILLC DS-2023-01: **Subhasree Patro**
Quantum Fine-Grained Complexity

ILLC DS-2023-02: **Arjan Cornelissen**
Quantum multivariate estimation and span program algorithms

ILLC DS-2023-03: **Robert Paßmann**
Logical Structure of Constructive Set Theories

ILLC DS-2023-04: **Samira Abnar**
Inductive Biases for Learning Natural Language

ILLC DS-2023-05: **Dean McHugh**
Causation and Modality: Models and Meanings

ILLC DS-2023-06: **Jialiang Yan**
Monotonicity in Intensional Contexts: Weakening and: Pragmatic Effects under Modals and Attitudes

ILLC DS-2023-07: **Yiyan Wang**
Collective Agency: From Philosophical and Logical Perspectives

ILLC DS-2023-08: **Lei Li**
Games, Boards and Play: A Logical Perspective

ILLC DS-2023-09: **Simon Rey**
Variations on Participatory Budgeting

ILLC DS-2023-10: **Mario Giulianelli**
Neural Models of Language Use: Studies of Language Comprehension and Production in Context

ILLC DS-2023-11: **Guillermo Menéndez Turata**
Cyclic Proof Systems for Modal Fixpoint Logics

ILLC DS-2023-12: **Ned J.H. Wontner**
Views From a Peak: Generalisations and Descriptive Set Theory

ILLC DS-2024-01: **Jan Rooduijn**
Fragments and Frame Classes: Towards a Uniform Proof Theory for Modal Fixed Point Logics

ILLC DS-2024-02: **Bas Cornelissen**
Measuring musics: Notes on modes, motifs, and melodies

ILLC DS-2024-03: **Nicola De Cao**
Entity Centric Neural Models for Natural Language Processing

ILLC DS-2024-04: **Ece Takmaz**
Visual and Linguistic Processes in Deep Neural Networks: A Cognitive Perspective

ILLC DS-2024-05: **Fatemeh Seifan**
Coalgebraic fixpoint logic Expressivity and completeness result

ILLC DS-2024-06: **Jana Sotáková**
Isogenies and Cryptography

ILLC DS-2024-07: **Marco Degano**
Indefinites and their values

ILLC DS-2024-08: **Philip Verduyn Lunel**
Quantum Position Verification: Loss-tolerant Protocols and Fundamental Limits

ILLC DS-2024-09: **Rene Allerstorfer**
Position-based Quantum Cryptography: From Theory towards Practice

ILLC DS-2024-10: **Willem Feijen**
Fast, Right, or Best? Algorithms for Practical Optimization Problems

ILLC DS-2024-11: **Daira Pinto Prieto**
Combining Uncertain Evidence: Logic and Complexity

ILLC DS-2024-12: **Yanlin Chen**
On Quantum Algorithms and Limitations for Convex Optimization and Lattice Problems

ILLC DS-2024-13: **Jaap Jumelet**
Finding Structure in Language Models

ILLC DS-2025-01: **Julian Chingoma**
On Proportionality in Complex Domains

ILLC DS-2025-02: **Dmitry Grinko**
Mixed Schur-Weyl duality in quantum information

ILLC DS-2025-03: **Rochelle Choenni**
Multilinguality and Multiculturalism: Towards more Effective and Inclusive Neural Language Models

ILLC DS-2025-04: **Aleksi Anttila**
Not Nothing: Nonemptiness in Team Semantics

ILLC DS-2025-05: **Niels M. P. Neumann**
Adaptive Quantum Computers: decoding and state preparation

ILLC DS-2025-06: **Alina Leidinger**
Towards Language Models that benefit us all: Studies on stereotypes, robustness, and values

ILLC DS-2025-07: **Zhi Zhang**
Advancing Vision and Language Models through Commonsense Knowledge, Efficient Adaptation and Transparency

ILLC DS-2025-08: **Sophie Klumper**
The Gap and the Gain: Improving the Approximate Mechanism Design Frontier in Constrained Environments

ILLC DS-2025-09: **Jordi Rudo Weggemans**
Quantum versus classical resources in computational complexity

ILLC DS-2026-01: **Bryan Eikema**
A Sampling-Based Exploration of Neural Text Generation Models

ILLC DS-2026-02: **Marten Folkertsma**
Empowering Quantum Computation with: Measurements, Catalysts, and Guiding States

ILLC DS-2026-03: **Valentin Richard**
Presuppositional and Dynamic Aspects of Questions

ILLC DS-2026-04: **Puyu Yang**
Bringing Science to the Public: The Role of Wikipedia in Scientific Communication

ILLC DS-2026-05: **Johannes Kloibhofer**
Cycles with Annotations: Non-Wellfounded Proof Theory of Modal Fixpoint Logics

ILLC DS-2026-06: **Danish Kashaev**
Approximation via duality in offline, online and strategic settings

ILLC DS-2026-07: **Oskar van der Wal**
Taking a Step Back: Measuring and Mitigating Bias in Language Models

ILLC DS-2026-08: **Lynn Engelberts**
Classical and Quantum Cryptanalysis of Lattices and Codes

ILLC DS-2026-09: **Xiaoyu Tong**
Dissecting Incongruity: Metaphor and Humor Understanding of Large Language Models

ILLC DS-2026-10: **Qian Chen**
Lattices of Tense Logics: Tabularity, Completeness and Decidability

ILLC DS-2026-11: **Daniil Khaitovich**
From arbitrary sets to social groups: group agency in multi-agent modal logic

THE ROLE OF UNDERSTANDING AND TIMING IN THE ETHICS OF QUANTUM TECHNOLOGIES

