LUND UNIVERSITY

Imagining the Impossible - Conflicting Norms and Values on Social Sustainability, Security, and Artificial Intelligence

Malmio, Irja

2025

Link to publication

Citation for published version (APA):
Malmio, I. (2025). *Imagining the Impossible - Conflicting Norms and Values on Social Sustainability, Security, and Artificial Intelligence*. Division of Risk Management and Societal Safety, Faculty of Engineering, Lund University.

Total number of authors:
1

Creative Commons License:
Unspecified

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Download date: 04. Jul. 2026

Imagining the Impossible

Conflicting Norms and Values on Social Sustainability, Security, and Artificial Intelligence

IRJA MALMIO
DIVISION OF RISK MANAGEMENT AND SOCIETAL SAFETY | LUND UNIVERSITY

This thesis examines the relationship between AI, security, and social sustainability as a sociotechnical system, exploring the implications and perspectives that emerge from this perspective. It represents a compilation of four articles that rely on both conceptual insights and empirical findings. The articles incorporate diverse perspectives to enhance the comprehension of AI's effects on social sustainability and security. The thesis concludes that the ethical dimensions of security and sustainability stem from foundational tensions, which balance universalist claims with competing objectives that promote value pluralism and diversity. These tensions are evident in AI imaginaries through various ethical processes that are deeply influenced by political and power-based motives. This thesis has established a conceptual groundwork for examining AI through a holistic lens of social sustainability and security. Additional contributions include expanding the conceptual understanding of security and identifying ethicalization processes in AI imaginaries. This research's practical aim is to inform policymaking while providing conceptual insights into navigating complex issues, such as AI regulation and its societal implications.

Printed by Media-Tryck, Lund 2025 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

Faculty of Engineering
Division of Risk Management and Societal Safety

ISBN 978-91-8104-499-7

Imagining the Impossible

Conflicting Norms and Values on Social Sustainability, Security, and Artificial Intelligence

Irja Malmio

LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Engineering (LTH) at Lund University to be publicly defended on 21st of May 2025 at 09.00 in Lecture Hall V:A, Division of Risk Management and Societal Safety, LTH

Faculty opponent
Professor Cecilia Åsberg, University of Linköping

Organization: Lund University, Faculty of Engineering,
Division of Risk Management and Societal Safety

Document name: Doctoral Dissertation

Date of Issue: 2025-04-23

Author(s): Irja Malmio

Sponsoring organization: Swedish Defence University

Title and subtitle:
Imagining the Impossible – Conflicting Norms and Values on Social Sustainability, Security, and Artificial Intelligence

Abstract:
This thesis examines the relationship between AI, security, and social sustainability as a sociotechnical system, exploring the implications and perspectives that emerge from this perspective. It represents a compilation of four articles that rely on both conceptual insights and empirical findings. The articles incorporate diverse perspectives to enhance the comprehension of AI's effects on social sustainability and security. The thesis concludes that the ethical dimensions of security and sustainability stem from foundational tensions, which balance universalist claims with competing objectives that promote value pluralism and diversity. These tensions are evident in AI imaginaries through various ethical processes that are deeply influenced by political and power-based motives. This thesis has established a conceptual groundwork for examining AI through a holistic lens of social sustainability and security. Additional contributions include expanding the conceptual understanding of security and identifying ethicalization processes in AI imaginaries. This research's practical aim is to inform policymaking while providing conceptual insights into navigating complex issues, such as AI regulation and its societal implications.

Key words:

Social Sustainability, Security, Artificial Intelligence, Sociotechnical Imaginaries, Values, Technology Development

Supplementary bibliographical information

Language English

Number of pages: 86

ISBN: 978-91-8104-499-7 (print)

ISBN: 978-91-8104-500-0 (pdf)

Recipient's notes | Price | Security classification

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature Date 2025-04-01

Imagining the Impossible

Conflicting Norms and Values on Social Sustainability, Security, and Artificial Intelligence

Irja Malmio

LUND
UNIVERSITY

Supervisor
Professor Misse Wester, Division of Risk Management and Societal Safety, Lund University

Co-supervisors
Professor Hans Liwång, Department of Systems Science for Defence and Security, Swedish Defence University

Senior Lecturer Petter Narby, Department of Systems Science for Defence and Security, Swedish Defence University

Faculty Opponent
Professor Cecilia Åsberg, University of Linköping

Examining Committee
Associate Professor Malin Gawell, University of Södertörn

Senior Researcher Henrik Carlsen, Stockholm Environment Institute

Associate Professor Jussi Karlgren, University of Helsinki

Funding organization
Swedish Defence University

Number of pages: 86

Coverphoto by Irja Malmio and Gemini
Copyright pp 1-86 Irja Malmio

Paper 1 © Irja Malmio and Sustainability Science Springer Nature
Paper 2 © Irja Malmio and Discover Sustainability Springer Nature
Paper 3 © Irja Malmio and Technology in Society Elsevier LTD
Paper 4 © Irja Malmio and Misse Wester (Manuscript unpublished)

Faculty of Engineering
Division of Risk Management and Societal Safety

ISBN 978-91-8104-499-7 (print)
ISBN 978-91-8104-500-0 (pdf)

Printed in Sweden by Media-Tryck, Lund University, Lund 2025

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

To my mother, Tanja, with all my love

Summary

Social sustainability and security are two expansive concepts that often contain conflicting elements while acting as reciprocal pairs, where one component necessitates the other. These ideas have been thoroughly debated since the introduction of Sustainable Development, as highlighted in the Brundtland Report from 1987. Simultaneously, the operation of AI systems worldwide and their impact on human conditions have emerged as crucial aspects of contemporary society to monitor and analyze, particularly from the perspective of social sustainability. This thesis, therefore, aims to establish and describe the link between AI, security, and social sustainability as a sociotechnical system while considering the implications and perspectives that emerge from such an understanding. Unraveling this link has the potential to foster technology development that alleviates societal challenges in alignment with ideals associated with social sustainability. Another benefit is identifying properties in these systems that lead to "sociotechnical harms," where bias can entail increased polarization, intolerance, and, ultimately, a deteriorating security situation.

This thesis is a compilation of four articles that base their findings on conceptual contributions and empirical findings. The articles draw from diverse perspectives to widen the understanding of AI's impacts on social sustainability and security. The thesis concludes that the ethical dimensions of security and sustainability stem from foundational tensions, balancing universalist claims with competing objectives that advocate for value pluralism and diversity. These tensions are manifested in AI imaginaries through various ethicalization processes that encompass political and power-based motives.

This thesis establishes a conceptual foundation for addressing AI through a holistic perspective on social sustainability and security. Further contributions include broadening the conceptual understanding of security and identifying ethicalization processes in AI imaginaries. The practical contribution of this research is directed to policymaking and offers conceptual insights on how to address complex issues, such as AI regulation and its societal impacts.

Keywords: Social Sustainability, Security, Artificial Intelligence, Sociotechnical Imaginaries, Values, Technology Development

Abbreviations

AI	Artificial Intelligence
AGI	Artificial General Intelligence
ANT	Actor-Network Theory
CSS	Critical Security Studies
IR	International Relations
LLMs	Large Language Models
SCOT	Social Construction of Technology
SDGs	Sustainable Development Goals
STI	Sociotechnical Imaginaries
STS	Science Technology Society
TA	Thematic Analysis
UN	United Nations
UNDP	United Nations Development Programme

Acknowledgments

This thesis process has been a journey of self-discovery and personal growth, but it would have meant nothing without the people who have supported me along the way. Therefore, I would like to take this opportunity to express my gratitude to those who have played a crucial role in this progress. First, I would like to extend a warm and heartfelt thank you to my supervisors for their guidance, expertise, and support throughout this journey. Misse Wester, you are undoubtedly the most fabulous supervisor, role model, and friend I could have asked for—thank you for the inspiring conversations about research, life, and everything in between. Hans Liwång, thank you for providing opportunities that broadened my horizons and for, during my moments of doubt, instilling in me your encouraging motto, "It's a feature, not a flaw." Petter Narby, thank you for helping me refine my conceptual ideas and for your patience in reading my drafts again and again. I am forever grateful for your intellectual support and insightful comments on my work. I would also like to extend my gratitude to Finn Nilsson and Henrik Hassel for their thoughtful comments on my thesis draft, which significantly contributed to improving it. I also want to thank Mathias Leese for his valuable comments on my half-time draft and for posing critical questions that helped me advance my work. Additionally, I would like to thank all the respondents who contributed with their time and insights to the data collected in Article IV.

I am deeply grateful to my academic friends, who have inspired and supported me throughout my journey. To my dear friends and mentors, Aida Alvinius and Arita Holmberg, thank you for your unwavering faith in my ability and for showing me the doors I needed to open to pursue my dreams. Aida, I don't think I would ever have been where I am today without your encouragement and support. Arita, your valuable input and support throughout my thesis journey have been indispensable. To my partners in academic hardships, Maja Svennbro, thank you for the fantastic discussions about everything from obscure French philosophers to deep analyses about why men go fishing. You are an essential part of my intellectual journey, and I look forward to more deep conversations about the invisible forces of power and progress. A very warm thank you to Frida Linehagen, Marcus Dansarie, Therese Tärnholm, Julie Decarpentrie, Emma Fredriksson, Henrik Paulsson, and Rasmus Andrén. Thank you for your camaraderie, encouragement, and shared moments of struggle and triumph.

I also extend my heartfelt thanks to all my colleagues at the Department of Systems Science for Defense and Security at the Swedish Defence University, as well as to my colleagues at the Division of Risk Management and Societal Safety at Lund University. I would especially like to thank Elin Darte Hasselvärn for your administrative support over the years and for always cheering me on. Evelyn Salas-Alfaro, you're next—I'm rooting for you! Therese Almblad, Astrid Sjölin, Adriana Ávila-Zúniga Nordfjeld, Johan Granholm, Nicholas Engelhardt, Suus Hopman,

Thomas Frisk, Eva Lagg, Martin Lundmark, Patrik Stensson, and Ingrid Kihlander, thank you for your valuable comments on my work and the many meaningful conversations we've shared over the years. Vera van der Zoest, thank you for assisting me with the tables in Article II and for your helpful advice on everything from handling journal reviews to career planning. Hanna Jungwalius, you are a wise woman! Thank you for your insightful feedback on several of my manuscripts and for providing me with critical insight into the military context. Kent Andersson, your positive energy and encouragement have been a constant source of motivation throughout this thesis work. A special thank you to Stefan Silfverskiöld—I will never forget your support during my first conference appearance; it meant the world to me.

I would also like to extend my gratitude to the people outside my department who have helped me throughout the years: To my first department at SEDU, Leadership Command and Control. Thank you: Sara Bondesson, Sofia Nilsson, Magna Robertsson, Maria Fors Brandebo, Eva Johansson, Claes Wallenius, Torbjörn Engelkjes, Helena Hermansson, Erik Hedlund, Daniel Hjalmarsson, and Pär Daléus. A very special thank you to Oskar Sjöström for your constant support and friendship and for challenging my cognitive skills in Wordfeud! A warm thank you to the administrative staff at FHS, especially Simon Ydhag, for your exceptional technical support and for helping me resolve issues that would have taken me weeks to figure out on my own in just a matter of seconds. Additionally, I want to thank Caroline Bjurström for cheering me on and resolving all my problems simply by typing an email.

In addition, I would like to acknowledge the research communities that have inspired me over the years: ISDRS International Sustainable Development Research Society, where I would especially like to thank Sebastian Thomas, Malin Gawell, Olga Cam, Simon Lockrey, and Peter Dobers, thank you for your valuable feedback on my work and for making me feel so welcomed among your peers with my out-of-the-box perspectives. I also want to thank Simon Hollis and the Climate, Crisis, and Security Network at the Swedish Defence University. Next on my list are all the great people I've encountered at AI Sweden; thank you for all the great conversations on AI, and I look forward to many more! I would also like to acknowledge the members of the NATO research group on Op Sec and Information Security—Keith, Andrew, Nina, Tineke, Therese, and Mills—thank you for broadening my horizons (especially on all the NATO acronyms) and for being very cool, smart, and fun people to hang around with. To all the fantastic research communities I've had the privilege of coming in contact with so far, thank you for the inspiration. I look forward to future collaborations!

A very special thank you to my dear friends—Stina Lantz, Björn Jonsson, Jenny Falk, Erika Howard, Ethel Linnel, Christina Sundin, Anna Frank, Fredrika Nordin Klinic, Tania Carson, Diana Vaivode, Sabina Helmola (who also helped me with tables and fancy illustrations for conferences!) and Sonja Dandenell—for your

constant moral support and encouragement throughout this journey. You are the best friends one could ever wish for, and your presence in my life surpasses what words can express.

To my beloved family, thank you for your endless love and support. To Katia Malmio and Per-Christian Sørlie, thank you for adding a silver lining to my life with fantastic dinners and parties! Igor and Nova, my beautiful children, being your mum is such a treasure; I love you to the moon and back. To my amazing mother, Tanja, thank you for always believing in me, no matter what crazy dream I've decided to pursue (well, you were a little skeptical of me becoming a ballerina). You have always been my most enthusiastic supporter and defender, always telling me in a prophetic voice, "Your time will come." Thank you for being the best mom anyone could wish for.

Finally, to Magnus Sahlgren, the love of my life, I cannot thank you enough for your boundless love, emotional support, and intellectual contributions to my thesis and my thoughts. You are my greatest source of inspiration, and I am so grateful for your beautiful energy and presence in my life. I love you forever.

List of Articles and Author Contributions

Article I

Malmio, I. (2024). Can Security be Sustainable? Three Perspectives on Security and Social Sustainability: Co-production, Paradox, and Deconstruction. *Sustainability Science.* 19(2) Doi:10.1007/s11625-023-01450-w

As the sole author of this article, I have engaged in all stages of the research independently.

Article II

Malmio, I. (2024). Artificial Intelligence and the Social Dimension of Sustainable Development – Through a Security Perspective. *Discover Sustainability* 5(1), 466. Doi: 10.1007/s43621-024-00677-6

As the sole author of this article, I have engaged in all stages of the research independently.

Article III

Malmio, I. (2023). Ethics as an Enabler and a Constraint – Narratives on Technology Development and Artificial Intelligence in Military Affairs through the Case of Project Maven. *Technology in Society.* 72 Doi: 10.1016/j.techsoc.2022.102193

As the sole author of this article, I have engaged in all stages of the research independently.

Article IV

Malmio, I. & Wester, M. (2025). Death by a Thousand Papercuts - A qualitative Study on Professionals' Imaginaries of AI, Social Impacts and Security. (Submitted)

As the first author of this article, I collaborated with my co-author on all stages of the research.

Table of Contents

1 Introduction

> It's all a question of imagination. Our responsibility begins with the power to imagine. It's just as Yeats said: In dreams begins responsibility. Turn this on its head, and you could say that where there's no power to imagine, no responsibility can arise.
>
> Haruki Murakami - *Kafka on the shore* (Murakami and Gabriel, 2005: 141)

We live in a time when competing demands for transformation and change intersect with multiple choices on how to achieve the best outcome, not only in the present but also in an unforeseen future. However, despite ample evidence that the world is in a dire state, budding aspirations for creating a sustainable society seem to have given rise to continuous disputes and uncertainties about how to best proceed. The year 2024 and its continuation in 2025 have proven particularly turbulent, as the Bulletin of Atomic Scientists announced the yearly predicted Doomsday Clock's standing at 89 seconds to midnight, which is the closest to a global catastrophe it has ever been (Mecklin, 2025). Reasons for this calamitous alarm have been explained by a deteriorated global security situation stemming from wars in Ukraine and Gaza combined with insufficient measures to reverse climate change and mitigate the consequences of global poverty and injustice (UN, 2022). In addition, the rapid progression in Artificial Intelligence (AI), especially in areas of Large Language Models (LLMs), image recognition, and visual reasoning (Maslej, et al. 2024), has created increasing concerns among international experts and policymakers about the social risks of AI (Bengio, 2023), with its potentially adversarial effects on democracy (Runciman, 2019) and global security (Johnson, 2019).

Furthermore, the extensive amount of information on how to create alternative outcomes appears to have led to a "comprehension deficit," where making decisions about the best outcomes has become increasingly complex (Smil, 2022). In conclusion, the issue of sustainable technology development has emerged as a wicked problem involving multiple stakeholders with competing agendas, unclear boundaries of jurisdictions and obligations, and a constantly changing landscape of tangled interdependencies concerning material and social resources, and consequently, no clear-cut solutions on the table (Weber & Khademian, 2010).

So, how can we address and comprehend these interconnected challenges? What theoretical tools are needed to unlock the black box of AI's future trajectories while fostering a technology development that is socially sustainable? One critical piece

in understanding this intricacy is identifying how the entanglement of imaginaries, political structures, values, and material networks affects sustainability and security measures. In other words, approaching the complexity that our sociotechnical world entails requires a thorough analysis of the ideas that drive various choices on real-world issues combined with the theoretical tools that facilitate the implementation of practical interventions (Walker, 1984).

However, pursuing research from this perspective opens up a montage of socially conflicting values, juridical requirements, political motives, the materiality of technology, the people involved in its design, maintenance, and production, and the sociotechnical system that encompasses them all. In this multifaceted mixture, technology and artifacts play a critical role in fulfilling societal functions (Geels, 2005), which means that the influence of technology is essential when analyzing social change (Gerhold & Brandes, 2021). As suggested by Hannah Arendt (1958):

> The objectivity of the world – its object- or thing-character – and the human condition supplement each other; because human existence is conditioned existence, it would be impossible without things, and things would be a heap of unrelated articles, a non-world if they were not the conditioners of human existence (Arendt, 1958: 9).

In other words, regardless of how the question of human progress is approached, its intertwinement with technological development is an iceberg that cannot be ignored. Therefore, there is a need to revisit the persistent question of technology's role in shaping the composition and evolution of society, as well as its interconnected impact on security, democratic institutions, and human perception.

1.1 Research Aim

This thesis aims to explore the relationship between social sustainability, artificial intelligence (AI), and security. Several academics, policymakers, and leading experts have identified the connection between security and sustainability as paramount in the quest for sustainable development for future generations (Acharya, 2001; Crabtree, 2020; Hanlon and Christie, 2016; King and Murray, 2001; Saetra, 2022). This perspective has laid out a theoretical foundation that permeates the work in the UN on sustainability and peace but has also received criticism for not being radical enough. The critics argue that its principles fortify an obsolete model of Western hegemony hidden in a cloak of humanitarianism (Chandler, 2008; Paris, 2001; Wibben, 2008) while firmly pushing through development narratives to pursue a capitalist agenda (Duffield, 2007; Edwards, 2021; Pulido & Peña, 1998). What tends to be missing in this discussion, however, is a sociotechnical approach that considers the implication of technology for humanity's direction by addressing

mutually reinforcing dynamics encompassing technologies, security aspects, policies, and social behaviors (Liwång et al., 2023; Savaget et al., 2019).

As of today, the technosphere—the aggregated mass of all technological artifacts on Earth—exceeds the total biomass of humans (Zalasiewicz et al., 2016). This has placed humanity in the era of the Anthropocene, where human activities are exerting significant impacts on the environment at all scales (Crutzen, 2006). In the current technological metamorphosis, AI plays a critical role as a transformative meta-technology and is involved in most technological appliances today (Suleyman and Bhaskar, 2023). The magnitude of AI's relevance to human progress has sparked an increasing interest in academia and the public, resulting in a doubling of AI publications since 2010, with many AI-driven technologies advancing scientific progress on a broad scale (Maslej et al., 2023). In 2024, two AI-related Nobel Prizes were awarded in physics to John J. Hopfield and Geoffrey E. Hinton for their inventions in artificial neural networks and to Demis Hassabis, John Jumper, and David Baker for their achievements in computational protein design (nobelprize.org). Additionally, the beginning of 2025 marked a significant disruption in the global market with the introduction of the Chinese LLM DeepSeek as a cheaper and more effective alternative to its American counterpart, ChatGPT, sparking substantial discussions on the technological race between the US and China and highlighting several geopolitical issues (Shaikh, 2025). Meanwhile, the Dhrago report, released in June 2024 by the European Commission, warned that while Europe struggles to balance demands for AI regulation with maintaining its innovation capacity, it may (or has already) become utterly irrelevant in the global AI arena (Dhrago, 2024). Thus, the entire geopolitical landscape, as well as the human condition, has been fundamentally reshaped at its core by the development of AI. Hence, to grasp this progression and understand where AI's future trajectories might lead, it is necessary to broaden the perspective to include interconnected elements of social phenomena, political aims, security concerns, conflicting imaginaries, and the power structures that follow.

Accordingly, this thesis has evolved around exploring how we can understand social sustainability, security, and AI as a sociotechnical system and what implications and perspectives arise from such an understanding. This aim emphasizes the relationship between security and social sustainability, defined by its openings and boundaries, synergies and trade-offs, and how this entangled connection is manifested in technology. In this regard, AI illustrates the interactions between the physical and the social, epitomizing central questions such as technological agency and human control (Leese, 2019; Zanzotto, 2019), value alignment (Gabriel, 2020; Ryan, 2024), social justice (Osoba et al., 2019), and broad societal questions about how to promote diversity (Moon, 2023), while minimizing the adverse effects of polarization (Burton, 2023). In the context of security, this development has also had profound effects, manifested in increasing cybersecurity awareness (Backman, 2023), ethically induced

debates over algorithmic policing (Egbert & Leese, 2021), and the rising occurrence of hybrid threats, such as information warfare (Straub, 2019).

1.2 Research Questions and Research Process

This thesis is based on four articles addressing the relationship between social sustainability, security, and AI, which are studied as integrated aspects of a sociotechnical system. In alignment with the research aim, this research has proceeded from two research questions that should be interpreted in light of the contextual and theoretical framework presented in the next chapter, which limits the possible answers to the otherwise broad research questions.

To lay out a theoretical foundation for this thesis, the first research question was aimed at investigating:

- Using a sociotechnical perspective, how can we address the conceptual relationship between social sustainability, security, and AI?

This question is investigated in Articles I and II. Article I is a conceptual paper that examines various normative and value-based frameworks influencing the relationship between social sustainability and security. The conceptual framework is inspired by an analytical approach to analyzing the normative and context-sensitive relationship between social sustainability and security, drawing on three theoretical perspectives: paradox, co-production, and deconstruction. Article II builds upon the theoretical framework of co-production presented in Article I by examining the sociotechnical connection between social sustainability, security, and AI, as well as its representation in academic literature. This article bridges into the second research question:

- How is the connection between social sustainability, security, and AI imagined across different contexts?

This research segment has been explored in Articles II, III, and IV. Article II examines the context of academic literature and policy reports, as well as their relationship to the interplay between social sustainability, security, and AI. Article III examines the context of civilian and military applications of AI through the media-induced debate that followed the public revelation of Google's involvement in Project Maven in 2018. Article IV examines the context of discursive environments among AI professionals in an interview study with thirteen AI experts who were interviewed about the potentials and risks associated with the development of AI and their views on the interconnection between social aspects and security.

Together with the research aim, the two research questions have contributed to building a theoretical and empirical body of knowledge on which this thesis draws its conclusions, which will be presented in Chapter 6: Conclusions.

1.3 Overview of the Thesis

The structure of this thesis starts with a chapter on the theoretical framework guided by the research aim described above. It examines prior research on social sustainability, security, and AI, providing the reader with the theoretical context from which this thesis departs. This chapter begins with a review of the connection (and disconnection) between social sustainability and security, which is further analyzed in Articles I and II. Next, a section on the academic field of science, technology, and society (STS) follows. STS is often described as challenging the utility-focused view of technology that views technology as separate from its social context. This section examines the reflexive interactions between humans and technology, emphasizing how the materialization of values is a crucial component of a sociotechnical system. Following this, a section on sociotechnical imaginaries (STI) is presented, which is a vital conceptualization for linking technology to social factors and informs the thesis' main theoretical framework. The final section of the theoretical background contextualizes AI and the various imaginaries associated with it, which are further elaborated in articles II, III, and IV. Thereafter, a chapter is presented on the epistemological and empirical methods that have guided this thesis, along with a discussion of the methodological considerations highlighted in the articles. Next, a summary of the papers is presented. The thesis synthesizes the main findings in Chapter 5, Discussion, highlighting social sustainability and security as a complex sociotechnical system, as well as the ethicalization of conflicting values, and what a co-production perspective can contribute to understanding the dynamic presented in this thesis. The thesis concludes with a summary that reviews the research aims and questions, highlighting its main contributions and limitations, and outlines potential avenues for future research to build upon this work.

1.4 Publications

I. Malmio, I. (2024). Can Security be Sustainable? Three Perspectives on Security and Social Sustainability: Co-production, Paradox, and Deconstruction. *Sustainability Science.* https://doi.org/10.1007/s11625-023-01450-w

II. Malmio, I. (2024). Artificial Intelligence and the Social Dimension of Sustainable Development – Through a Security Perspective. *Discover Sustainability* 5, 466 (2024). https://doi.org/10.1007/s43621-024-00677-6

III. Malmio, I. (2023). Ethics as an Enabler and a Constraint – Narratives on Technology Development and Artificial Intelligence in Military Affairs through the Case of Project Maven. *Technology in Society.* https://doi.org/10.1016/j.techsoc.2022.102193

IV. Malmio, I., and Wester, M. (2025). Death by a Thousand Papercuts - A Qualitative Study on Professionals' Imaginaries of AI, Social Impacts, and Security (Submitted)

2. Theoretical Framework

This section provides context for the theoretical framework from which this thesis departs, beginning with the conceptual pair of social sustainability and security. Following this, attention shifts to the academic field that views science, technology, and society (STS) as a dynamic ecosystem composed of social values, material artifacts, human agents, politics, organizational stakeholders, and institutions. In this setting, a central element is sociotechnical imaginaries (STI), used as a conceptual tool in this thesis to connect technology's societal role to security aspects influenced by dimensions of values and power. Lastly, a section contextualizes AI and its associated imaginaries.

2.1 Social Sustainability and Security

Sustainable development is a wide-ranging concept that embodies both purely instrumental qualities and a more comprehensive approach, shedding light on power asymmetries, resource allocation, and aspects of inclusion and diversity (Malmio and Liwång, 2022). Accordingly, sustainable development proceeds from a holistic assessment of three building blocks: ecological, economic, and social sustainability (Elkington, 2008), thus embracing a transdisciplinary and holistic outlook that acknowledges the relevance of social and cultural needs to achieve a sustainable future (Savaget et al., 2019). In this aspect, social sustainability plays a crucial role and appears particularly prominent when examining crises and security issues, as many areas prioritized to maintain a secure society are also identified as goals for social sustainability. This sentiment is expressed in the UN's 17 Sustainable Development Goals (SDGs) formulated in Agenda 2030 and adopted by the UN Assembly in 2015, where five of the 17 goals are specified as social goals, including Goal 1: No poverty, Goal 4: Quality education, Goal 5: Gender equality, Goal 10: Reduced inequalities, and Goal 16: Peace, justice, and strong institutions (UN, 2015).

However, a universal definition of social goals related to sustainable development has not been adopted. Instead, many discipline-specific descriptions have focused on various aspects depending on the context in which they appear (Dempsey et al., 2011; Litting & Griesler, 2005; McKenzie, 2004). Several studies have focused on the context of urban development, where suggested topics for the definition of social

sustainability are community resilience (Dempsey et al., 2011; Magis, 2010), social cohesion (Al-Homoud & Tassinary, 2004), and inclusion and exclusion (Haddon, 2000). In addition, social sustainability has been explained in terms of societal metabolism, which aims at conceptualizing the connections between symbolic and material systems (Littig, 2002). Amartya Sen (2004) has proposed a policy-oriented approach that equates social goals with capabilities, shifting the focus from allocating material resources to how people can express their agency. Accordingly, these capabilities refer to basic human needs, such as "the ability to move about, to be clothed and sheltered, and the power to participate in the social life of the community" (Sen 2004: 78). Attempts have also been made to propose a functional definition to suit all contexts. Ballet et al. (2020) suggest this definition should encompass social cohesion, equity, and economic security. In addition, social sustainability has also appeared in several related areas, such as social entrepreneurship, focusing on innovations that cater to the needs of society (Gawell, 2013), social justice, which focuses on distributive questions of the allocation of social and material resources (Moroni, 2019) and critical sustainability studies, which address the social and environmental consequences of unequal power structures in society (Rose & Cachelin, 2018).

Social sustainability and the connection to security encompass conflicting elements attributed to the highly normative influence of the contested concept of "security" (Smith, 2005). The traditional description of security proceeds from ideas associated with political realism, which views politics as governed by objective laws, takes a demonological approach to the notion of the antagonist, is motivated by the accumulation of power as its central goal, and justifies acts of war as a moralistic concern (Morgenthau & Thompson, 1993). From this perspective, security is narrowly understood as various objectives of individuals and states aimed at protecting and preventing future attacks from antagonistic threats (Rothschild, 1995). A crucial aspect is achieving a "strategic power balance," where technology plays a central role as a force multiplier. Henry Kissinger (1961) famously described this act as a product of three variables: "power, the will to use it, and the assessment of these by the potential aggressor" (Kissinger 1961: 12).

The conceptual foundations of "security" have been thoroughly explored by scholars associated with Critical Security Studies (CSS), who challenge the conventional problem description in International Relations (IR) (Peoples & Vaughan-Williams, 2020). These critics argue that focusing solely on the problem-solving dimension of world politics risks neglecting other crucial elements. Instead, we need to consider components in our social world that construct security and insecurities while meticulously observing how the concept of security alters our sense of what should be prioritized and what is ignored. Arguably, conceptual understandings of security embody negative or positive framings (Hoogensen Gjorv, 2012) that can have a narrow focus on locating and neutralizing a threat

(Hartman et al., 2012) or a broad focus that includes bodily, communal, productive, national, and financial aspects (Nelson, 2024).

A broadened perception of security highlights the explicit and implicit power dimensions manifested in security measures as relational and constructed (Huysmans, 1998), while the problem formulation itself becomes a study area of inherent political and value-based motives (Taureck, 2006). Consequently, attaching "security" to something suggests a deliberate power move, paving the way for exceptional measures to protect the object under threat (Buzan et al., 1998) while legitimizing the securitized object's hidden nature from public examination (Ericson and Wester, 2022). When the concept of security is deconstructed in this way, it opens up critical questions, such as who is the subject of security (Butler, 2009), who exhibits agency to securitize (Spivak, 2015), and what can be considered a security issue (Walker, 2016), thereby challenging the taken-for-granted narratives of victims, perpetrators, and saviors (Wibben, 2011).

Consequently, the critical approach to security has been aimed at reformulating the referent object of security, shifting its focus from the state's survival to the needs of the individual. Ideas that prioritize the individual as the central focus of security while highlighting "emancipation" as the key to achieving "true" security (Booth, 1991) have been particularly influential in shaping the UN's global approach to peace and sustainability. Emancipation has thus contributed to a holistic view of security that encompasses social and economic components and is pivotal in "human security," the theoretical cornerstone of the UN Human Development Report, released by the UN Development Programme in 1994. The report expresses a close kinship between security and the social dimension of sustainability to build sustainable peace and accomplish "freedom from fear and want" (UNDP, 1994: 24). There is thus a strong association between emancipation and development, which connects insecurity and conflict with underdevelopment while supporting the idea that "sustained, inclusive, and sustainable economic growth is essential for prosperity" (UN General Assembly, 2015: §9). Accordingly, resourceful states should protect and improve the situation for people living in low-income countries. By providing this help, they will benefit themselves by making the world a more secure place (Duffield, 2007).

Another approach to connecting security with social aspects can be found in the term "societal safety," which has been gradually integrated into the Scandinavian safety vocabulary since its introduction in the late 1990s, linking the concept of security to various risk-assessment methods (Aven, 2009). The concept of societal safety has been mostly connected with the area of crisis management with a strong focus on undesired events prevention, building trust in societal functions, and governance structure for restoring society and environmental conservation (Høyland, 2018). The concept is closely associated with maintaining critical societal functions, protecting citizens' lives and health, and meeting their basic needs during stressful situations such as crises, armed conflicts, and pandemics (Olsen et al.,

2007). The emphasis on safeguarding essential societal functions operates at different levels: the national, regional, or local municipality (Hassel et al., 2022).

However, a critique of the attempts to reformulate security in general, and human security in particular, is that they fail to address the "techno-materiality of security" (Müller and Richmond, 2023), which can be explained as the strategic practice of designing or utilizing technology to achieve security and/or political objectives (Hecht, 2011). Even if this connection does not seem far-fetched, it has, according to Daniel McCarthy (2017), been largely absent in conventional security studies. While the influence of technology has been recognized at a surface level, the more profound effects of techno-materiality on world politics in shaping and maintaining hybrid forms of power have not been sufficiently addressed (McCarthy, 2017).

Nonetheless, the sociotechnical systems that arise from such arrangements tend to reflect and reproduce the hierarchies and inequalities present in the societal structures. To capture this sociotechnical arrangement, social sustainability has been selected as an analytical lens in this thesis because it offers a broader array of analytical possibilities while avoiding the pitfalls associated with conceptualizing human security, such as Western hegemony and the inadequate consideration of technology for understanding international relations.

Accordingly, from this brief description of the different ways of approaching security and social sustainability, further discussed in Article I and II, a leading guideline in this thesis is understanding security as a relational and normative concept, which can be understood in narrow (focused on threat reduction) or broadened terms (focused on the security of individuals). However, even when the security concept is broadened, power is an integral dimension that needs to be acknowledged. Social sustainability, too, is a normative and context-dependent concept, and, as Michael Redclift (2005) puts it, "sustainable development is necessary for all of us, but it may be defined differently in terms of each and every culture" (Redclift, 2005: 213). Thus, defining social sustainability within the theoretical scope of this thesis proceeds from the goals expressed in Agenda 2030 as institutional justice, social equity, and inclusion (UN General Assembly, 2015), as these are relevant aspects to develop from a security perspective while pronouncing sociotechnical components as a "common good" (Radder, 2019). This leads us to the next section, exploring sociotechnical systems as conceptual tools.

2.2 Sociotechnical Systems, Values, and Uncertainty

There are many ways of approaching technology as a research object (Mitcham, 1994), and selecting the parameters that define its conceptualization will also impact the research outcome (Hansson, 2015). The word "technology" originates from the Greek words *tekhne* (skilled craft) and *logos* (speaking or reasoning), implying the systematic study of a technique (Acemoglu and Johnson, 2023). Tekhne was initially perceived as the "practical wisdom" of tools and machines and disclosed as a distinct type of skill from other types of knowledge, such as philosophy and mathematics (Hansson, 2015). Accordingly, technology was identified as "applied science," signifying a distinct area of expertise while establishing science as a more refined type of knowledge (Houkes, 2009; Mitcham, 1994; Radder, 2019). The early association between technology and practical function has contributed to a view of technology as inanimate objects evolving in deterministic patterns, associating technology development with an unstoppable momentum that exists independently of its social context (Jasanoff, 2016). In addition, technological artifacts, in their deep association with mechanical production, are closely linked with the scientific ideal of normative objectivity (Douglas, 2009; Feenberg, 2017; Kroes, 2012). Andrew Feenberg (2002) describes the objectivist position as a positivist approach to "modeling neutrality by instrumental means" (Feenberg, 2002: 6). Accordingly, the traditional view on technology is based on a detachment from value-based motives, indicating that the ultimate purpose of technological development is increased efficiency, while advancements in other variables are subsidiary to that goal.

However, a sharp demarcation between, and hierarchical classification of, technology and science, as well as between facts and values, have proven challenging to maintain due to an increasing overlap among science, social values, politics, and technology. Accordingly, the reluctance to accept technology, science, and social values as separate fields of knowledge has opened up a vast area of research on social studies in science, technology, and society, STS (Latour, 2004; MacKenzie, 1993; Pinch & Bijker, 1984; Wajcman, 2000; Winner, 1980). This academic field encompasses varied research orientations but proceeds from the understanding that technology and knowledge production comprise a diverse and hybrid composition of technological and social elements. Furthermore, it recognizes the interactions between humans and machines as creating "mutual intelligibility" (Suchman 2007: 34) and, therefore, challenges the longstanding distinction between the physical and the social spheres. Hence, technology is not simply the application of new methods to produce material goods. It concerns everything involved in its production, including visioning, imagining, and the politically motivated choices ultimately manifested in the artifact itself (Acemoglu and Johnson, 2023).

The academic approaches to defining the composition and dynamics of a sociotechnical system are varied, covering multi-perspective approaches at the

intersection of sociology, institutional theory, and innovation studies (Geels, 2004). One significant influence is social constructivism, which emphasizes that meaning and social reality are contextualized subjective experiences (Guzzini, 2000). Thus, the focus is shifted from explaining how technology functions in a societal context to understanding the social dynamics involved in *interpreting* this function, which can produce various outcomes depending on its social context (Pinch & Bijker, 1984). This conceptualization is associated with the social construction of technology (SCOT), stressing that societal influence exerted by technology is ontologically constituted within a particular context. Therefore, it is not a fixed property but can vary depending on geographic location and socioeconomic factors, such as gender and age (Manjikian, 2018). However, as Stefano Guzzini (2000) has pointed out, a shortcoming in constructivist theories is the eclectic and redundant tendency to produce "so what" conclusions. To address this shortcoming, it is crucial to focus on comprehensive empirical studies that explain how narratives of politics, power, technology, and transformation are constructed and how they influence society. In this regard, actor-network theory (ANT) offers an empirical lens for understanding how the world is constructed and stratified (Latour, 2004) through the agency of non-human actors, the 'missing masses' (Latour, 1992). Technologies stabilize and enable human practices, but in doing so, they also constrain and condition them, creating a socio-material system of domination and resistance (Bueger & Stockbruegger, 2018). These questions have also been addressed in feminist technoscience studies, which focus on understanding the dynamics involved in the co-creation of technology and various social categories in forming hegemonic structures (Åsberg & Lykke, 2010). The gendering of technological artifacts proceeds from "what is most obvious, therefore most hidden" (Feenberg and Beira 2018: 40) – values that are transmitted in the design processes of technological systems and solidified in artifacts, therefore enabling value judgments, possible biases, and power relations that often remain undetected (Faulkner, 2001; Haraway, 1991; Harding, 1986; Michelfelder et al., 2017; Simon, 2016; Wajcman, 2009). In this critical view of sociotechnical systems, the circumferences of the metaphysical concept of "human" are challenged, as well as what it means to be an embodied human subject in a materialized world (Åsberg & Braidotti, 2018; Åsberg & Lykke, 2010; Hayles, 1999; Loh, 2019).

Although these theoretical contributions emphasize different aspects of a sociotechnical system, they share similar traits. Firstly, technology is not neutral but intertwined with political motives (Winner, 1980, 2020). This assumption invites an analysis of technology that proceeds from understanding the political and social shaping of technology, stressing reciprocity between social values and technological development (Brey, 2018; MacKenzie & Wajcman, 1999). Since political motives are value-based (Winner, 2020), paying attention to how values operate in a sociotechnical system can "provide the possibility for drawing links between individual, social, structural, and cultural levels of analysis" (Hitlin and Piliavin 2004: 383).

Values can be defined as collectively held conceptions that define "good" and "bad" conduct, establish ideal forms of behavior, assign priority to different points of view, and influence the preferences of individuals, organizations, and society (Chong, 2000). Values possess both motivational and normative qualities (Hitlin & Piliavin, 2004) and are deeply intertwined with our sense of identity (Korsgaard, 1996). Values, therefore, have fundamental implications for how human beings perceive and are being perceived as subjects within a society (Rosa & Trejo-Mathys, 2013). They are often resistant to change and can be a fertile ground for conflict and disagreement (Leese, 2017). Therefore, legislators and decision-makers must handle value-based issues with care, as overlooking this dimension can lead to a legitimacy crisis (Svedin, 2011). However, there is no single correct way of conceptualizing social values, as different conceptualizations offer limited perspectives. According to Kenter et al. (2019), this calls for addressing "complex, wicked problems, where facts are uncertain, stakes are high, and decisions are urgent" (Kenter et al., 2019: 1440). Furthermore, choices about developing new systems are inherently normative, as all problem descriptions are partially shaped by the value lenses through which they are viewed and formulated. Therefore, different perspectives and lenses can produce competing ontological views, a topic investigated in article I. In a sociotechnical context, values have a formative effect that affects all parts of technological innovation, procurement, and usage (Friedman, 1997). This influence is especially significant when examining the dimensions of technology and security, where the values attributed to how "security" is perceived can have considerable consequences for public acceptance of security technologies (Gerhold & Brandes, 2021). This, in turn, can lead to increased militarization (Enloe, 2016) and erode democratic resources (Amoore & Raley, 2016) while fueling uncertainty and fostering mistrust (Callon, 2009).

Secondly, while social aspects are an integral part of a sociotechnical approach, there is similarly a call for a "material turn" in political science, which understands the materiality of things and their performative effect (Butler, 1990) as profoundly intertwined with security politics (Amoore, 2016; Aradau et al., 2015; Müller & Richmond, 2023; Shaw, 2016). One example of this approach is the work of Karen Barad, who examines the assemblage of social and technological components through the lens of quantum physics, where elementary particles do not have an absolute form but can appear as a manifestation of "wave-particle duality" (Barad, 2007). Hence, physical objects do more than impact social outcomes; they *intra-act* with discursive practices and produce specific security practices, leading to hierarchizations and exclusions (Aradau, 2010). In this context, Airport security can be used as an illustration of how technological devices are associated with security imaginaries (Law and Singleton, 2000). Through compulsory transparency, these technologies provide a *feeling of security* while simultaneously turning their users into "willing suspects" (Hall, 2015). In this way, airport security devices constitute a material-discursive performance of security (Schneier, 2009) that excludes

differentiation and creates the presence of "the other," a potential terrorist (Los, 2004).

A third characteristic of a sociotechnical system is that because it comprises several parts where every element is susceptible to independent variation (Bennett, 2010), it continuously produces a certain amount of uncertainty. Uncertainty is created from the entangled characteristics between mutually reinforcing dynamics between technologies, policies, and social behaviors (Savaget et al. 2019), further reinforced by the inherent framework and contextual setting (Sollie, 2007). These propensities make it challenging to map out unintended consequences or predict positive spillover effects (Suleyman & Bhaskar, 2023) and change unsustainable characteristics that appear (Meadows, 1997). Pierre-Benoît Joly (2015) addresses the uncertainty in a sociotechnical system as a complex movement of the "liquid society," where science and technology constantly evolve through *a state of permanent change*, making the trajectories of emerging technologies challenging to forecast and modify (Joly, 2015). Another factor contributing to uncertainty in sociotechnical systems is the dimension of time. Ray Kurzweil (2004) talks about the "law of accelerating returns," relating to feedback loops that increase technological advances in unpredictable ways (Kurzweil, 2004). Accordingly, the novelty and rapid pace of emerging technologies, such as AI and robotics, present vast and intractable uncertainties about their benefits, risks, and future directions (Wallach & Marchant, 2019). Hence, the question of how to address the high degree of uncertainty in sociotechnical systems is a significant concern for the experts of AI technology in knowledge-based system research (Voorbraak, 1996).

However, predicting emerging technology developments is a precarious task, as explained in the Collingridge dilemma. As technology evolves and can be shaped, its societal impacts remain uncertain. However, when technology becomes societally embedded and its implications are known, it becomes difficult to change an undesirable development (Collingridge, 1980). This kind of dilemma is similar in ethical evaluations, where the ethics of technology always seem to be either "too early"—evaluating technologies without knowing their direction—or "too late"—understanding the ethical impact of technology but at a moment when the technology is less prone to change (Kudina & Verbeek, 2018).

In summary, there are numerous ways to approach technology as a research subject. This thesis has adopted a sociotechnical perspective that acknowledges the interplay among various components shaping the sociotechnical system. A crucial element of this perspective is the visionary aspect of technology development, which combines organizational aspects of path dependency and visioning with elements of values, power, and uncertainty. This leads us to the next section on sociotechnical imaginaries and their impact on the evolution of technology.

2.3 Imaginaries of Technology, Security, and Power

The idea that technology and social aspects are interconnected aligns with concepts that relate material elements to the human imagination (Borup et al., 2006). Jasanoff and Kim refer to this aspect as "sociotechnical imaginaries" (STI), which are "collectively held, institutionally stabilized, and publicly performed visions of desirable futures" (Jasanoff and Kim, 2009: 120), combining a normative approach to imagination with the materiality of networks (Jasanoff, 2015). These imaginaries are influenced by make-believe perspectives and anticipatory practices (Roßmann, 2021) and contain dystopic and utopic components (Jasanoff, 2015) that influence worldviews, ideals, production, and consumption systems (Hagbert et al., 2020). STI reflect and reinforce social norms embedded in specific institutional contexts, shaped by individual sentiments transferred to technology (Sartori & Theodorou, 2022; Suchman, 2007). As such, STI can be approached as modern myths that blend technical and scientific possibilities to justify fictional ideas that reinforce various political standpoints (Wall, 2008). Narratives and technological storytelling are essential as fundamental animators of STI (Cave, Dihal, and Dillon, 2020). However, while narratives and discourses focus on language, ideas, and social practices, STI emphasizes purposiveness, action, and aspiration in technological materializations (Beck et al., 2021).

For an STI to be successful, it must demonstrate performative power achieved through repeated integration into policies, strategies, technologies, and societal practices (Miller, 2020). Another feature of imaginaries is that they are inherently oriented toward future trajectories, thus favoring scientific and technological change while simultaneously reinforcing existing structures (Borup et al., 2006). Combining current knowledge with a future-oriented perspective provides the actors involved with emotional and evaluative significance, functioning as a coping strategy to navigate uncertainty in sociotechnical systems (Adams, 2023). The future is uncertain; imaginaries emphasize that uncertainty, using visions of technological innovation and progress to overcome it (Adloff and Neckel, 2019; Miller, 2020). Furthermore, an evolving sociotechnical imaginary requires a counter-imaginary. The future is a realm of possibilities, characterized by varied tensions and parallel dynamics that portray the world as a process and a becoming. Therefore, future imaginaries can be oppositional and antagonistic, yet they also generate new possibilities (Ahlqvist, 2022). In this way, STIs are a vehicle for the co-production and co-evolution of technology and society (Jasanoff, 2015).

Sociotechnical imaginaries rely on visioning and persuasion power (Acemoglu and Johnson, 2023) and often appear as taken-for-granted assumptions embedded in the viewpoints of political elites (Kruck and Spencer, 2013). Accordingly, power is an essential aspect of imaginaries that reconfigures actors' sense of the rightness of action, their sense of possible spaces and forms of action, and their agency, or lack thereof, in the future. STI, therefore, serve as modes of societal self-organization

that contribute to opening up or closing down possible horizons of future action (Stirling, 2007). In this way, STI set the parameters for political actors to make decisions or by making particular choices more or less plausible. Notably, this dimension of power is inextricably linked to resistance, which entails a productive element of power that simultaneously enables and constrains (Allen, 1999; Peoples & Vaughan-Williams, 2020). Power forms a complementary relation to resistance, and *where there is power, there is resistance* since individuals contest fixed identities and relations in ongoing and sometimes subtle ways (Deveaux, 1994: 231). In this way, imaginaries assert their influence in fluctuating and unsettled manners, serving as vehicles through which power is distributed (Foucault, 1997). A similar expression for this dynamic is "soft power," which operates through a strategic application of imaginaries linked to democracy, human rights, and sustainable development, thus creating influence through a humanitarian appeal rather than coercion (Nye, 2004). Hence, the imaginary of sustainable futures carries an ideational power that co-opts rather than forces, persuades rather than dominates, and includes rather than destroys .

The conceptualization of imaginaries has created an open, contested, and dynamic field influenced by many discursive and intellectual approaches (Bächle and Bareis, 2022; Jasanoff, 2015). Examples of disciplines where imaginaries have been particularly influential are innovation studies (Konrad & Böhle, 2019; Meyer, 2019), sustainability research (Adloff & Neckel, 2019; Delanty, 2021), political science (Beck et al., 2021) and war studies (Müller & Richmond, 2023; Öberg, 2019).

In the realm of security, a convergence has emerged between scientific advancements in technology and war, with a significant impact in the Western liberal world. The hybridization of technoscience and security has contributed to imaginaries that co-evolve around the overarching themes of "antagonism" (Lawson, 2011), "ethical war" (Zehfuss, 2018), and "precision and accuracy" (Suchman, 2020). One common trait in these imaginaries is that they are influenced by ethicalization processes, where ethical arguments are used to justify security measures (Rychnovská, 2016). One example is the imaginary of war as a "humanitarian project" constructed by ethics, policing, and householding (Öberg, 2019), which has profoundly impacted how the overall security agenda in the UN is shaped and consequently acted upon (Shepherd, 2021). However, ethicalization processes have also become increasingly apparent in public discourses surrounding the use of AI, which is further discussed in Article III.

The imaginary of war as justified by ethical motives, fortifies understanding the act of killing as "scientific warfare" (Bousquet, 2022), and forges a strong connection between the scientific developments of technological innovation and the security domain (Ford, 2017). Technological progress in the security domain is strongly linked to the notion that improved accuracy results in fewer casualties and has been a leading paradigm in military strategy dating back to the Cold War era (Suchman,

2020). However, when probed more closely, "accuracy" is a highly normative concept that has evolved in a complex process of conflict and collaboration between various social actors, including technicians, politicians, military leaders, laboratories, and the organization itself. From this perspective, "accuracy" as a military imaginary has been a shaping force but has also, in turn, been shaped by its context (MacKenzie, 1993). According to Maja Zehfuss (2018), there is an instrumental dimension in framing accuracy as an ethical construction. Ethics, she suggests, is formed around a paradox that functions both as an enabler and sets a boundary for the military agenda since "ethical war is not just a solution to perceived problems; it is also a practice that is shaped by the vision that drives it" (Zehfuss, 2018: 50).

While technology and security are intimately linked with ethical imaginaries, there is also a growing tendency toward a "technization of security" (Gerhold & Brandes, 2021), where technical systems increasingly fulfill society's security demands and expectations. This trend is particularly evident in the rise of AI-driven security solutions, such as surveillance algorithms, predictive policing systems, and biometric identification tools. A strong belief in technological solutions to societal problems stems from the assumption that machines are more effective than humans at performing specific tasks—such as processing vast amounts of data, identifying patterns, and making decisions with speed and precision (Osoba et al., 2019). However, once these security devices are implemented, it is challenging to reverse the obstruction of the democratic liberties they were initially intended to protect. This brings us to the final part of the theoretical framework: the impact of AI on security and social sustainability.

2.4 AI in Context

AI traces its roots to Alan Turing's groundbreaking article "Computing Machinery and Intelligence," published in 1950 in the journal Mind (Turing, 1950), and was formally established as an academic discipline at the Dartmouth Conference in 1956 (Leins, 2019). Since then, AI has experienced several cycles of scholarly and public interest, with a notable surge in engagement following the release of ChatGPT in November 2022 (Coeckelbergh & Gunkel, 2023).

When examining AI as a research topic, there are as many approaches as there are questions to ask, problems to solve, and perspectives to explore, encompassing vast areas such as robot ethics (Gunkel, 2012), engineering (Vaswani et al., 2017), neuroscience (Hassabis et al., 2017), social approaches to digitalization (Cheney-Lippold, 2017), religion (Singler, 2019), and many more. However, in general terms, AI can be described as an attempt to replicate the workings of human intelligence by creating a self-learning technology (Collins, 2021). Through its

practical application, a range of definitions of AI has emerged, highlighting various capabilities associated with its development (Wooldridge, 2021). In addition, the "black box" operations related to AI have also contributed to building an air of mystique, making it a challenging research object (Straube, 2019). In this spectrum, AI has become closely associated with technological storytelling and science fiction, which has obscured the boundaries of what AI can do in its current state and its future potential (Cave et al., 2020; Padden & Öjehag-Pettersson, 2021; Sartori & Theodorou, 2022). Narratives of the potential capabilities associated with AI often appear in a dichotomous vocabulary revolving around an impending AI doom (Tegmark, 2018) and a "post-scarcity utopia" (Boström, 2024). Thus, a significant gap exists between AI's purely technical functions and the socially constructed imaginaries associated with the technology.

When focusing on the technological function, AI can be described as a machine-based system that infers and generates outputs such as predictions, content, recommendations, or decisions, using different levels of autonomy after deployment (OECD, 2023). AI is based on algorithms, which can be a sequence of simple instructions or describe complex mathematical equations (Gillespie, 2014). Algorithms operate as a single unit or perform more complicated tasks in a network interacting with other technological systems (Horowitz, 2018). In this way, AI technology possesses a generative capability that can accelerate the pace and complexity of technological innovation (Schmidt, 2023), leading some experts to argue for the singularity hypothesis —a point at which AI becomes "super intelligent" (Kurzweil, 2004). Additionally, AI is a general-purpose technology with significant implications for the geopolitical landscape (Verdegem, 2024). Suleyman and Bhaskar (2023) use the definition "transformative meta-technology," denoting AI as a ubiquitous technological system, which "is itself a maker of tools and platforms, not just a system but a generator of systems of any kind" (Suleyman and Bhaskar, 2023: 78). AI is typically categorized into artificial general intelligence (AGI), which signifies automation of the full spectrum of human intellectual capabilities, and narrow AI, which concentrates on the performance of specific tasks (Wooldridge, 2021).

At an ontological level, AI is closely related to algorithmic inference methods and mathematical reasoning. Therefore, AI has been rendered with the imaginary of "superior objectivity" (Anichini & Kotras, 2024) while simultaneously being discussed as a transmitter and producer of social norms (Beer, 2009; Burrell, 2016; Pasquale, 2015). Thus, a discrepancy exists in the academic literature regarding AI, its association with mathematical precision, and its tendency to generate and exacerbate social bias. Mathematical reasoning was strongly favored by Alan Turing, who believed that mathematical conclusions present knowledge in a way that "cannot be seriously doubted" (Turing, 1939: 216). This line of reasoning aligns with the belief that mathematical reasoning constitutes the ultimate form of

objective knowledge, a claim that has been reinforced throughout the history of science (DeWitt, 2018).

However, this idea relies on the premise that information only moves in one direction, thereby ignoring the reflexive nature of any knowledge production, where the observers are intertwined with the system being observed. This tendency is particularly prominent in AI systems, making AI's contribution to knowledge production a meta-epistemological issue, where data formed by a complex social reality are condensed into an output, which is then mirrored back into the system. Another aspect of this epistemic co-production is the condensation of a complex social reality into a singular mathematical output, which is claimed to represent the entire sample. Thoughts like these have been particularly influential in critical algorithmic studies (Seaver, 2019), where scholars have highlighted the mismatch between the social world and its algorithmic representations, which can lead to adverse societal effects (Amoore, 2020; Aradau, 2004; Beer, 2009; Bengio, 2023; Coeckelbergh, 2023; Collins, 2021; Pasquale, 2015). The critics claim that the association between AI and mathematical reasoning creates a false impression that algorithmic inference methods are free from bias, when, in reality, the influence of bias cannot be cleared (Amoore, 2020; Beer, 2017; Mau & Howe, 2019; Pasquale, 2015).

Social bias is perhaps not a new phenomenon, but its extent in automated processes can amplify existing inequalities (Browne, 2015; Bucher, 2018; Eubanks, 2017; Orike & Ene, 2023; Sartori & Theodorou, 2022). Blindness to bias in the data can be problematic in automated decision-making, particularly in areas where social sorting occurs, such as policing (Egbert & Leese, 2021), judicial settings (Završnik, 2020), and financial and insurance decisions (Brenner & Hartl, 2021). Reliance on AI in official decision-making has been referred to as "algorithmic governmentality" (Rouvroy and Berns, 2013), indicating an increasing reliance on algorithms to guide human actions and validate information rather than trusting human authority (Lustig et al., 2016). However, the processes from which a particular decision was made often remain opaque, providing a limited understanding of the underlying mechanisms that prompted a specific outcome (Burrell, 2016; Mau & Howe, 2019; Pasquale, 2015). The effects of digital monitoring frequently have a more negative impact on marginalized groups in society and can increase the persecution of minorities (Browne, 2015; Eubanks, 2017). Another problem is the lack of high-quality data, with negative consequences on healthcare, especially for women and minority groups who are underrepresented in medical data (Goh & Vinuesa, 2021).

Despite extensive discussions of "objectivity" and bias in the societal applications of AI, its association with mathematical precision and accuracy has been highly sought after by the security domain (Amoore & Raley, 2016; Beer, 2017; Suchman, 2020). Accuracy delivered by algorithms enables faster and more reliable application of complex information flows to support cyber-attacking software,

surveillance, and tracking systems (Crawford, 2021; Elliott, 2018), capabilities that could ultimately save lives. However, AI's security-related (and other high-risk) applications have raised ethical concerns about autonomy and levels of human control. The capability to learn from previous outcomes without human intervention (Scott, 2021) enables AI systems to act autonomously to a greater or lesser extent (Leese, 2019; Scharre, 2018). Autonomy is, therefore, a central feature of AI, referring to the levels of capability a machine has to act independently of human interference. Autonomy depends on the type of task the machine is performing, the relationship of the human to the machine when performing that task, and the level of sophistication of the machine's decision-making when performing the task (Scharre, 2018). Fully autonomous AI systems in the military context raise concerns about the use of lethal force without human involvement (Leese, 2019). Considering that nuclear delivery systems could operate autonomously, this presents a daunting prospect (Johnson, 2019). Consequently, autonomy issues frequently correspond with questions of transparency and explainability, as well as who or what should be held accountable when things go wrong (Macnish, 2018). In this context, human-in-the-loop is a critical concept that refers to the levels of human control required to ensure that humans make the final decision (Zanzotto, 2019).

Another perspective of AI and algorithmic networks is to view them as "mediums of power" (Cockburn, 1985), signifying a multitude of interlaced systems influencing humanity in various ways (Crawford, 2021). A fundamental aspect of this configuration involves control of material assets, including ownership of data (Hummel et al., 2021), access to critical minerals such as lithium and cobalt (Kalantzakos, 2020), resource management (Walia et al., 2024), and the manifestation of values (Birhane et al., 2022; Mohamed et al., 2020). One critical parameter is access to data, which lies at the heart of AI innovation, manifested through a growing field of data commercialization (Couldry & Mejias, 2019), data mining (Harrag & Alshehri, 2023), and datafication, where most aspects of human existence are transformed into data (Sadowski, 2019).

Data as a commodity highlights the inherent ambiguities and paradoxes regarding material and socio-cultural ownership (Hummel et al., 2021), exemplified by the inherent conflict between protecting users' autonomy and privacy while considering them as data-producing assets (Paltieli, 2022). In addition, the industrial landscape of AI is dominated by *Big Tech*, a handful of companies with the financial resources necessary to develop AI (Verdegem, 2024). American institutions produce most of the world's significant language and multimodal models (Ligett et al., 2024). However, the American lead has been challenged by increasing Chinese competitiveness, and the introduction of DeepSeek in January 2025 caused considerable turmoil in the international tech scene (Shaikh, 2025). Still, the fact that the majority of LLM manufacturers are based in either the USA or China has sparked significant debates around AI sovereignty, particularly in the EU, which has contributed to a spiraling global AI race logic (Mügge, 2024). Either way, the fact

that the resources necessary to develop AI are uneven from a global perspective raises uncertainties about whose values are represented (Mohamed et al., 2020), highlighting issues of societal resilience (Bourbeau, 2015) and the security politics of dual-use technology (Martins & Ahmad, 2020). There are also concerns that the asymmetrical access to resources driving the cutting-edge development of AI could propel global inequality into an AI divide (Gehl Sampath, 2021; Goralski & Tan, 2022; Saetra, 2021).

In summary, AI as a research topic intersects with queries on the proliferation of power, epitomizing conflicts between safeguarding national perspectives and embracing universal values, tensions between open-source initiatives and capitalist principles, and the balancing act of responsible containment that does not stifle innovation.

3. Methodology

This section will discuss the philosophical and methodological approaches that have guided this thesis work and develop methodological considerations raised in the appended articles. I will begin with an overview of the philosophical positioning that influences the research, including critical theory, interpretive methods and reflexivity, and complexity as a research approach. Next, I will proceed with a section that describes the methods used throughout this thesis work.

3.1 Philosophical Positioning

Critical Theory

The philosophical foundation of this thesis draws on a critical theory approach to examine social sustainability, security, and AI as a sociotechnical system. Thus, a key aspect of this research is to explore what it means to adopt a critical approach and how this perspective shapes the research questions and their outcomes. Critical theory represents a loosely associated group of disciplines that seek to uncover the underlying cultural assumptions that dominate a field of study and the broader society (Mohamed et al., 2020). Thus, critical theory challenges the taken-for-granted assumptions that underpin traditional views on scientific knowledge, particularly positivism and empiricism (Smith, 2021), which uphold the value-free ideal and the epistemic value of objectivity as the pinnacle of science (Douglas, 2009; Elliott, 2017). A positivist stance on knowledge production views knowledge as an external reality (DeWitt, 2018) that is accessed by applying measurements to build empirical data (Desmet, 2022). However, this assumption has drawn criticism from a diverse array of scholars, who maintain that "objectivity" is both unattainable and unethical, a sentiment famously captured by Donna Haraway (1988) as "the god trick of seeing everything from nowhere" (Haraway, 1988: 581). A critical approach, therefore, challenges the assumption that the world consists of specific, determinate, and relatively identifiable processes (Law, 2004). As such, it aims to dismantle the "fortress," constituted by a "presentation of a defined problem, the development of linear argumentation for the exegesis of claim, and the anticipation and preemption of potential critique" (Ravecca and Dauphinee, 2018: 127). Whereas traditional views on science are rooted in a problem-solving paradigm, critical theory seeks to question the motives behind commonly held ideas about how

the world functions. The idea is that by widening the scope of analysis, the researcher will be able to spot essential dimensions that do not fit into a worldview solely focused on finding rational conclusions (Amoore, 2009; Aradau, Huysmans, Neal, & Voelkner, 2014; Feenberg, 2002; Gunkel, 2012; Law, 2004; Marcuse, 1968; Peoples, 2011; Walker, 2016; Young, 1990). In the words of Iris Young (1990):

> Critical theory denies that social theory must accede to the given. Social description and explanation must be critical, aiming to evaluate the given in normative terms. Without such a critical stance, many questions about what occurs in a society and why, who benefits, and who is harmed will not be asked, and social theory is liable to reaffirm and reify the given social reality (Young, 1990: 5).

However, as pointed out by Karen Barad in an interview with Juelskjær and Schwennesen in 2012, critical theory also has its shortcomings. The notion of critique is prone to reproducing a particular spatiality and temporality, which can lead to fortifying binary categories while situating critique from an outside perspective (Juelskjær & Schwennesen, 2012). Another issue, frequently pointed out as the big ogre in science philosophy, is the tendency to produce relativistic theories that are detrimental to scientific knowledge (Jackson, 2016). The question of relativism and truth has been debated at length throughout the history of science, but John Law (1990) argues that this need not be the disastrous scenario it has been portrayed as, given that the rules of the method differ across various contexts:

> The either/or thinking committed to absolutism is a false dichotomy. Instead, embracing epistemological relativism invites an essential form of intellectual caution: the sense that all knowledge is shaped and contingent and could be otherwise in some other world (Law, 1990: 6).

Feminist scholars have sought to address the issue of relativism by emphasizing that objectivity in critical research can be attained through reflexivity and contextualization, that is, by recognizing different standpoints that produce various knowledge claims (Harding, 1991). Accordingly, contextualized research acknowledges diverse perspectives by highlighting how different approaches emerge in specific locations to tackle particular challenges in a given context (Wibben, 2011).

Furthermore, in its quest to uncover ambiguity and complexity, another aspect of critical theory is that it often produces "messy research." By broadening the problem formulation instead of seeking specific solutions, it establishes a depiction of the world that can be incoherent, indecisive, and challenging to capture with a single theory. However, as John Law (2004) has suggested, the attempt to shape the complex and elusive reality into neat, compartmentalized categories misses the point. Instead, attuning to the mess can reveal opportunities to explore multiple realities and conflicts that remain concealed within existing discourses and practices (Law, 2004; Squire, 2013). Another suggestion is to approach messy research from

a process-oriented perspective that embraces a pluralistic value system where individual agency emerges as a focal point of development within a specific context of social practices (Nayak & Chia, 2011). Hence, at an ontological level, a critical view of science and the world, with its connected phenomena, acknowledges that "what exists are not things made, but things in the making" (James, 1909: 263).

For this thesis, a critical perspective has guided the formulation of the research questions and the analysis of the resulting data. Thus, critical theory has been used as a lens to identify the processes that shape existing imaginaries related to social sustainability, security, and AI, as well as to understand their influence on policy outcomes and societal structures. Another vital contribution of critical theory to this thesis is its expanded scope for addressing questions often overlooked in conventional views, which lies at the heart of this thesis' philosophical approach. Additionally, a critical approach is utilized to deconstruct concepts by examining inconsistencies and tensions within the overarching framework and its individual components.

Interpretative research and reflexivity

Interpretation is a fundamental part of research that highlights questions surrounding the division between objectivity and the researchers' emotional investment (Dauphinee, 2015), as well as the selection of sources to be used for analysis (Dauphinee, 2013; Fujii, 2010; Park-Kang, 2015; Zehfuss, 2007). Accordingly, interpretation as a methodological approach is an essential building block when (de)constructing ideas and imaginaries of how security, sustainability, and technology are related. This thesis utilizes an interpretative approach as a conceptual tool to decode the narratives and imaginaries associated with AI, security, and social sustainability as a sociotechnical system, offering explanations and justifications for both intended and unintended actions.

Interpretation stems from a hermeneutical tradition that is ontological rather than methodological (Wibben, 2011). As a philosophical concept, it relies on the idea that pure methods cannot extract the truth. What we perceive as *knowledge* is not a true reflection of an independent reality but a social production of a historically situated culture (Polkinghorne, 2000). However, this form of heuristic probing into contextualized knowledge raises questions about whether science can provide the tools necessary to differentiate between better and worse truths or even to define what interpretation essentially is. As an example of a positivist stance on this matter, Max Weber (2011) has suggested that because interpretation is linked to value judgments, it entails a practical evaluation that lies beyond the scope of scientific proposals. Consequently, even if separating empirical statements of fact from value judgments is challenging, it is better to "let the facts speak for themselves" (Weber, 2011: 10).

In contrast, a social-constructivist perspective on interpretation suggests that since human experience of the world is inherently a subjective process, value judgments are unavoidable (DeWitt, 2018). However, this view comes with inevitable consequences. By accepting that human behavior is shaped by its context, the focus shifts from observation to explaining the context that gives meaning to a specific behavior. Since the context can be endlessly varied, it requires a large quantity of background knowledge, and making an inventory of such knowledge often remains an ad-hoc procedure (Suchman, 2007). Consequently, interpreting experience is necessarily both subjective and susceptible to cultural constraints (Bal & Boheemen, 2017). Another issue with this stance, as noted by Maurice Merleau-Ponty (2014), is that interpretation can lead to "high-altitude thinking" and create a distance between the interpreter and the observed phenomena. Interpretation should, therefore, strive for a unification of human experience and the lived embodiment of that experience (the observed object) (Merleau-Ponty & Landers, 2014). This can be achieved by incorporating a reflexive practice into the research process that acknowledges the researcher's subjective standpoint while posing a significant question: *How does the author take responsibility for their subjectivity?*

Reflexivity implies that the researcher approaches the question of investigation from a standpoint of responsibility for the knowledge claims they produce in interpreting the social existence of others. In this process, the involvement of normative, political, and epistemological factors that inform specific knowledge claims should be clarified (Ramazanoglu & Holland, 2002). However, reflexivity is not just about being completely transparent with one's own state of emotions and opinions throughout the research process, it also acknowledges the co-constituted nature of the researcher and the research process itself by focusing on the ambiguities, dissonances, and differences of the multiple interpretations that emerge during this process (Wilkinson, 2013).

In a reflexive research approach, the belief that supposedly knowledgeable subjects exist fully formed before encountering their object of analysis is considered a fallacy, just as the idea that the subject and object are separate, distinct entities that enable a rational and objective analytical perspective. In reality, both are co-constitutive and active participants in the production of knowledge. Likewise, the discourse surrounding the research project reveals a bias rooted in historical, cultural, and social conditions that shape both its subjects and objects. Therefore, knowers are influenced by their context, where adherence to disciplinary knowledge and assumptions, along with their preconceived notions of what a discipline encompasses, shapes their perspective on new knowledge claims (Bonditti et al., 2014). Aradau et al. (2015) suggest that when approaching reflexivity in the context of security, it should be expanded to include an analysis of the effects that methods, as practices, have. To achieve this goal, the researcher must consider the political dynamics, strategic imperatives, and institutional facilitators that favor specific methods over others since methods are practices intertwined with power relations,

both in exercising and being shaped by those relations (Aradau, et al. 2014: 11). In the same line of reasoning, Lisa Wedeen (2010) addresses epistemological reflexivity towards debates about security more widely by:

> Posing questions about what bounds the discipline and normalizes its modes of inquiry, rendering other possibilities unsayable, unthinkable, irrelevant or absurd (Wedeen, 2010: 264).

As I was raised in a privileged part of the world, I have been careful not to assume a "Western" worldview is universal for everyone. This is a critical issue, particularly in relation to social sustainability and the UN's agenda on sustainable development, which aims to address questions of diversity while also considering the *people* as a unified entity (Telleria, 2021). Another issue that has necessitated reflexive action, arising from an interpretative methodology, is the imprecise boundaries between facts and fiction (Edkins, 2013; Park-Kang, 2015) or even how to provide a transparent methodology for describing interpretation itself (Fujii, 2010; Vastapuu, 2018). These have been ongoing questions in this thesis work, as in all research projects. Interpretation as a methodological tool thus entails responsibility for the interpretation being made. This is especially crucial when discussing the societal impacts of AI, as those outside the professional AI field rely on narratives that mediate between the technology world and the public sphere to understand the types of technologies being developed. In this regard, the researcher must be highly aware of the imaginaries reinforced through their own interpretation (Dillon & Schaffer-Goddard, 2023), as the outcome can significantly influence public acceptance or rejection of the AI system (Cave et al., 2020). To address these issues, I have adopted a reflexive approach to become aware of the types of imaginaries or narratives I contribute to through this research. This has been an ongoing process throughout this work, where I have had to remind myself that while I cannot escape my thoughts, I can critically engage with them as I do with the research I conduct.

Complexity as a research approach

This thesis adopts a holistic and multidisciplinary approach to addressing the complexity of interactions among humans, technology, values, and politics that a sociotechnical analysis reveals. The characteristic of this methodological approach is that it does not fit into current academic categories but instead offers new modes of analysis that are both complex and multifaceted (Law and Singleton, 2005). Addressing research questions from a platform of complexity progresses from the foundational *why* question focusing on research design to produce the knowledge required to achieve that objective, the *what* (Tobi & Kampen, 2018). Accordingly, these research practices proceed from the problems at the center of research over discipline-specific concerns, theories, and methods (Leavy, 2011). When research gravitates from the problem and its context, looking beyond disciplinary boundaries becomes necessary to construct novel methodologies required for addressing the

issue (Wickson et al., 2006). A (mono)disciplinary approach, in contrast, is prone to reducing a phenomenon to its components and thus becomes too limited to address complex questions (Tobi & Kampen, 2018).

Furthermore, complexity research is well-suited for addressing emergent properties, where a holistic perspective enhances the understanding of the composition of the investigated phenomena as a whole while also adding depth to each of its components. Since variation does not occur linearly but often happens in ways that do not involve just one possible outcome, this perspective offers flexibility throughout the research process (Byrne, 1998). This is particularly useful in the context of research at the intersection of (in)security, science, and technology, which requires researchers to become attuned to trouble (Bellanova et al., 2020: 87) and create intelligibility from the disorganization that complexity invites.

Complexity in research can be addressed through different approaches, typically defined as multidisciplinary, interdisciplinary, and transdisciplinary, with considerable confusion regarding the differences between them. One distinction offered by Roderick Lawrence (2010) is that multidisciplinary approaches juxtapose various disciplinary contributions, while interdisciplinary approaches are more coordinated and integrated. Transdisciplinary approaches are often driven by public needs and integrate contributions from multiple disciplines to build an expanded systemic framework that incorporates various disciplinary and interdisciplinary inputs (Lawrence, 2010). The idea behind all these approaches is that by transcending disciplinary borders, new research pathways are opened, and with that, new knowledge-building practices emerge (Leavy, 2011).

The research questions of this thesis have created a complex space that necessitates crossing disciplinary borders. To gain a broader understanding of the cross-fertilization among different perspectives in shaping imaginaries of security, social sustainability, and AI, a methodology has been developed to capture complexity while maintaining intelligibility without sacrificing too many finer details. Thus, this thesis has expanded its disciplinary belonging inspired by a "remix approach," which combines various disciplines, contexts, and areas through a fragmented methodology designed to develop specific research interests (Navas, Gallagher, and Burrough, 2018). In addition to providing researchers with disciplinary flexibility, it also highlights the impact of the disciplinary context on researchers' preconceived notions of that discipline, shaping their perspective on the knowledge claims that follow (Bonditti et al., 2015).

The strength of this methodological application lies in its ability to provide a broadened analysis of complex systems. Furthermore, it provides a way of relating to the macro and micro aspects of human experience without being aggregative or reductionist (Byrne, 1998). However, all manifestations of complexity cannot appear simultaneously. Therefore, the researcher is compelled to select, thereby imposing what Karen Barad (2007) refers to as an "agential cut", in which the

subject is distinguished from the object within the phenomena created through intra-action. In Barad's (2007) words:

> Since different agential cuts materialize different phenomena – different marks on bodies – our intra-actions do not merely effect what we know and therefore demand an ethics of knowing; rather, our intra-actions contribute to the differential mattering of the world. Objectivity means being accountable for marks on bodies, that is, specific materializations in their differential mattering (Barad, 2007: 178).

Thus, selecting from a vast universe of possibilities requires the researcher to take responsibility not only for the research that materialized but also for the parts that did not.

Philosophical positioning in summary

This thesis has established its philosophical positioning within a critical theory context through its formulation of problems and the application of deconstructive approaches. It has adopted an interpretive stance to the research area, drawing inspiration from a social constructivist perspective on knowledge production. However, as noted above, this perspective has shortcomings, particularly in producing relativistic scientific conclusions. This tendency has been addressed by employing a processual and contextualized methodology that emphasizes reflexivity throughout the whole research process. Furthermore, this thesis has advanced its research focus through a complexity approach, integrating insights from various disciplines and methods to present a holistic view of the investigated area. For this thesis, it has meant that the development of a research strategy has been driven by a problem-centered approach in designing the research topics and purpose (Leavy, 2011), thereby paying more attention to the investigated topic rather than adhering to the confinements of disciplinary traditions.

3.2 Methods and Materials

The articles in this thesis have employed a mixed-methods research design, incorporating both qualitative and quantitative approaches. Paper I is a conceptual paper that utilizes document analysis to establish a theoretical foundation in alignment with the research aim of this thesis. Articles II, III, and IV are empirical studies that cover various aspects highlighted in the conceptual framework.

An overview of methods and materials is listed in Table 1.

Table 1. Methods and Materials

Article	Method	Material
I	Conceptual Research Document Analysis	Policy documents
II	Scoping Study Thematic Content Analysis	Structured review of 62 research articles from WoS and SCOPUS and 11 reports
III	Case Study Narrative Analysis	Newspaper articles, policy documents, and web pages
IV	Interview Study Imaginaries Analysis	Interviews with 13 respondents

The methods used in each appended paper have served different but complementary purposes. For example, conceptual research and document analyses were employed to define the research areas, while empirical studies and policy documents were used to examine the research gaps.

Article I - Conceptual Research and Document Analysis

Article I has mainly been conducted as conceptual research. Conceptual methods typically employ an iterative process, building conclusions based on description and explanation, thereby achieving a better balance between theory-building and theory-testing research (Meredith, 1993). Accordingly, conceptual research primarily focuses on developing ideas rather than expanding on specific theoretical frameworks (Mhurchú & Shindo, 2016). Article I proceeds as an investigation of the relationship between "contested concepts," defined by Gallie (1955) as based on internally complex value achievements, which are variously describable and flexible. One good example of a contested concept is "security," which is strongly associated with a normative valance that makes agreeing on a single definition complex (Collier, Hidalgo, and Maciuceanu, 2006). Therefore, these concepts tend to invite pluralistic understandings when applied within a contextual framework. However, even though a concept has multiple interpretations, there is a tendency to understand an idea in a single and uniform way, defined by Sahlgren and Carlsson (2021) as the "singleton fallacy." Definitions are typically organized in a binary position, even though language can be understood in many ways. Therefore, one conclusion is that concepts are not neutral instruments, and different definitions afford different kinds of legitimation (Shotter, 1993). In addition, analyzing and deconstructing concepts highlights the value-based position that informs various policy directions and is thus an essential tool for understanding the cultural significance of concrete historical configurations (Weber, 2011).

In Article I, the analysis focuses on studying five UN policy documents that address the connection between sustainability and security. The analyzed material explores the connection between sustainability and security, employing different approaches

that reflect the specific contexts in which they were created, and provides a comprehensive account of how security and social sustainability have been discussed throughout the evolution of Sustainable Development. Two of them, "Our Common Future," also known as "The Brundtland Report," released in 1987 (Brundtland, 1987), and "Transforming our World: the 2030 Agenda for Sustainable Development," released by the UN General Assembly in September 2015 (UN 2015), are considered canonical documents in the UN work on Sustainable Development (Mensah, 2019) while presenting valuable insights on how security has been approached from a sustainability perspective. Three reports were included from the UNDP: "The Human Development Report," released in 1994 (UNDP, 1994), the first report in which the concept of "human security" is introduced; and "Human Security Now," also known as the Ogata-Sen report, from 2003 (CHS, 2003). These two reports are vital documents in the UN's formulation of Human Security and have been frequently discussed in academic literature. A more recent publication, "2022 Special Report on New Threats to Human Security in the Anthropocene: Demanding Greater Solidarity" (UNDP, 2022), provides an updated account of how security and its linkages to social sustainability are conceptualized today. In addition, relevant academic contributions and grey literature from security studies, sustainability, and human security have been included to exemplify the divergent standpoints generated by the theoretical perspectives of paradox theory, co-production, and deconstruction. The literature discussed has been applied to illustrate how the relationship between security and social sustainability is altered depending on which perspective is applied. Therefore, a limitation of this study is the textual body on which it has based its conclusion. However, the focus has been on analyzing the contrasting outcomes produced by distinct ideological vantage points rather than providing an exhaustive literature review.

The selected material was analyzed using three theoretical perspectives — paradox, co-production, and deconstruction — to identify how the meaning and effects of the conceptual pair of social sustainability and security are altered depending on the perspective applied. The paradox perspective stresses an essentialist view of values that complicates a reconciliation of social sustainability and security, as exemplified in the concept of "national security." In contrast, a co-production perspective proceeds from a constructivist position, which emphasizes reciprocity and co-creation, characterized by a broadened scope of "human security." The third analytical lens deconstruction offers an alternative approach by viewing the conceptual pair from a poststructuralist perspective, focusing on the underlying processes that produce meaning while paying attention to the hierarchical positioning of values.

Article II - Scoping Study and Thematic Analysis

Article II uses a scoping study to map the field of social sustainability, AI, and security. This study uses quantitative and qualitative methods to identify themes and

analyze academic discussions on social sustainability, artificial intelligence, and security. Scoping reviews allow the incorporation of various types of literature in the analysis corpus. Therefore, they are valuable tools for identifying more specific questions that can be addressed by a more precise systematic review (Munn et al., 2018), especially when the literature is complex and heterogeneous (Peters et al., 2020). The study was guided by three objectives: (1) to document the link between social sustainability, artificial intelligence, and security; (2) to identify the opportunities and obstacles of AI to either support or hinder the progress of social goals in Agenda 2030 from the perspective of security and (3), to categorize and describe how this connection is framed in the literature. The documents chosen for analysis were extracted from the Web of Science (WoS) and Scopus databases, which are widely recognized for their high-quality international academic publications (Fosso Wamba et al., 2021). First, a search strategy was drafted by compiling a list of key search terms and establishing inclusion and exclusion criteria (Sarrami-Foroushani et al., 2015). The publication interval was set between 2017 and 2024, with a final date of September 30. The chosen time interval indicates that the research avenue on the link between AI and the SDGs is a relatively recent area of interest, with the overall number of academic publications on AI-related topics steadily increasing since 2017 (Maslej et al., 2023). Based on four search strings detailed in Article II, the initial search identified 2204 documents: 124 were selected for further examination. Exclusion criteria in this selection included energy, food security, smart cities, education, health care, bioeconomy, and circular economy. Inclusion criteria included ethics, security, social good, human rights, and democracy. In addition, book chapters and conference papers were removed. The 124 documents were then carefully read, resulting in a final selection of 62 articles. The selection criteria at the final stage were that the articles had to mention AI and the social goals of the SDGs, which could be related to security. In addition, eleven reports were included in the analysis. These were qualitatively selected based on relevance to the study's purpose and referenced in the selected articles. The final analysis included 62 articles and 11 reports.

After the initial selection, the selected articles and reports were studied in detail and labeled according to the principles of thematic content analysis, which involves interpreting and developing themes rather than merely verifying earlier theoretical conclusions (Clarke & Braun, 2017). The selection of themes aimed to illustrate the central findings on how the literature discusses security-related issues, following the broad application of the concept of "security" as conceptualized by the UN (UN General Assembly, 2015; United Nations, 1994). Accordingly, the analysis identified six themes in the literature corpus: AI for social good, economic security, development and humanitarian aid, values, sociotechnical harms, and societal security.

Article III - Case Study and Narrative Analysis

Article III is based on a case study of the debate surrounding Project Maven from its public announcement in 2017 to its aftermath in 2018. Case studies offer a holistic and in-depth investigation, suitable for multi-perspectival analyses, which focus on the perspectives of specific actors and how their interactions unfold (Tellis, 1997). The ethical narratives presented in this article draw on empirical examples derived from news articles and reports accessible online, as well as official documents from various websites. The Google News feed engine was utilized to identify grey literature, starting with a broad search using the term "Project Maven" during the 2017-2018 timeframe, which yielded 3,950 hits. To find articles reflecting the differing views of military and civilian perspectives, a second search was conducted using the terms "Project Maven" and "Civil-Military Divide," resulting in 17 hits. From these, a selection was made based on the criteria that a document clearly expressed an ethical narrative regarding either civilian or military stakeholders or as part of a response chain, thereby revealing how the opposing sides addressed the narratives created by their counterparts. A narrower search was also executed to identify the official narratives presented to the public, primarily by the Defense Innovation Board and Google, focusing on government documents, blog posts, and opinion pieces. Additional material was included to provide background information and elucidate the strong stance taken by protesting Google employees, primarily from sources such as Wareham and the Campaign to Stop Killer Robots (Human Rights Watch, 2020), as well as opinion letters addressed to the United Nations, signed by various prominent tech industry entrepreneurs. In total, the dataset analyzed in this article included 25 articles, six reports, and eight blog posts.

The choice of documents for analysis follows a heuristic approach that narratology invites (Bal & Boheemen, 2017), illustrating how key players in this field formulated specific ideas that significantly influenced the ensuing debate. The selection was theory-driven, meaning that the choice of data was influenced by conceptual emergence and relevance rather than being derived from a preconceived theoretical framework (Holton & Walsh, 2017).

The material analyzed in this article employs broad themes to organize the analysis and identify significant discourses within each theme through narrative interpretation. A narrative approach engages with stories that have become accepted or dominant over time. Consequently, narratives encompass multiple political possibilities, which this approach seeks to illuminate (Shepherd, 2021). To support this objective, Article III employs a thematic narrative analysis to examine how civilian and military actors perceive a technology-induced security issue and the potential consequences of these perceptions.

Article IV – Interview Study and Thematic Analysis of Sociotechnical Imaginaries

Article IV is an interview study focusing on the imaginaries of thirteen AI experts and their perspectives on the connection between AI, social sustainability, and

security. This article examines professionals' perspectives on the potential benefits and risks of AI development at a societal level, employing a qualitative interview study combined with thematic analysis. The data collected consisted of semi-structured interviews with 13 professionals from various AI-related areas, representing academia, private companies, and NGOs. The respondents were identified through contacts, and some were identified through the informants' suggestions. Initial contact with the respondents was made via email. The interviews took place from December 2023 to January 2024 in Stockholm, Sweden, and followed an interview guide with open-ended questions, which is included in Appendix 1. The interviews were conducted in person and via Zoom and recorded using a Dictaphone. The interviews took approximately one hour and were conducted in English and Swedish. Prior to the interviews, the respondents were provided with information on the study's purpose and informed of the ethical guidelines, including their right to anonymity and the option to withdraw their consent to participate at any time (Elliott, 2017).

The interviews were transcribed verbatim and analyzed by both authors. In the initial reading, similar expressions and meanings were identified and clustered together according to the principles of thematic analysis (TA). The aim of TA is not simply to summarize the data content but to identify and interpret key features of the data, guided by the research question (Clarke & Braun, 2017). Four themes—reality, dystopia, regulation, and utopia—were identified during this process. The interviews were studied in greater depth during the second phase, and quotes were categorized into sub-categories. The quotes have been translated from Swedish to English, and purely colloquial expressions have been removed.

A limitation of the chosen research approach is the classic one of qualitative case studies, which is that extrapolation and generalizability are problematic (Bryman, 2015). Additionally, the interview participants were primarily from Sweden; therefore, the answers apply to a Swedish context. The interviewees were also selected as representatives for their organizations, and this could have prevented them from speaking freely. This specifically applies to the respondents belonging to security organizations.

Ethical considerations

Ethical considerations have been incorporated into the methodological design of this thesis in several ways. One essential component is carefully deliberating over how various values inform not only occurrences at a global policy level but are also incorporated into every aspect of human activity. Article IV used interviews to collect data. Since the interview questions did not contain any sensitive information, ethical approval was not applicable. To align the data collection with the ethical principles described by Bryman, which include the requirements of sound science – information, consent, confidentiality, and usage – the study respondents were provided with thorough information on the study's purpose (Bryman, 2015). The

respondents in this study were anonymized, and their organizational belonging was pseudonymized. In addition, the interview guide and the actual interviews were conducted using a reflexive approach, a mode of analysis that acknowledges that all questions inherently make exclusive decisions about what materializes and what does not (Graham, 2020; Gunkel, 2012). A reflexive approach is instrumentally valuable as it fosters greater self-awareness and self-reflection among both producers and consumers of such knowledge. Reflexivity thus aligns with the notion that validating a knowledge claim begins with theorizing the social conditions of its production (Jackson, 2016).

4. Summary of Articles I- IV

4.1 Article I

Can Security be Sustainable? Three Perspectives on Security and Social Sustainability: Co-production, Paradox, and Deconstruction

Security and sustainability are prioritized goals in the Western liberal world. Maintaining democratic resources while simultaneously strengthening society's ability to address security issues closely aligns with the ideals expressed in social sustainability. However, merging normative theories, such as security and social sustainability, creates conceptual difficulties that are challenging to resolve. This article examines five policy documents that provide a comprehensive account of how security and social sustainability have been discussed throughout the evolution of Sustainable Development; "Our Common Future," also known as "The Brundtland Report," released in 1987 (Brundtland, 1987), "Transforming our World: the 2030 Agenda for Sustainable Development," released by the UN General Assembly in September 2015, (UN 2015). Three reports were included from the UNDP: "The Human Development Report," released in 1994 (UNDP, 1994), "Human Security Now," also known as the Ogata-Sen report, from 2003 (CHS, 2003), and "2022 Special Report on New Threats to Human Security in the Anthropocene: Demanding Greater Solidarity" (UNDP, 2022). These documents were analyzed using three perspectives that address the discrepancies between security and sustainability from different ontological baselines: paradox, co-production, and deconstruction. The paradox perspective highlights the inherently divergent qualities of sustainability and security, suggesting a trade-off situation. Conversely, the co-production perspective views social sustainability as a vital component in addressing security issues, while security, in turn, is a prerequisite for sustainability. A third perspective, deconstruction, presents an alternative view on the polarized understanding of security and sustainability by acknowledging the undecidabilities and aporias inherent in normative concepts. The article concludes that a conceptualization of security and social sustainability should adopt a process-oriented and contextualized approach, where normative tensions offer crucial insights into how competing values manifest.

This article focuses on addressing a foundational aspect of this thesis: how social sustainability and security can be approached through different ideological lenses.

4.2 Article II

Artificial Intelligence and the Social Dimension of Sustainable Development – Through a Security Perspective

This article examines Artificial Intelligence (AI) and the social dimension of sustainable development from a security perspective, highlighting recurring themes regarding the potential of AI to either support or undermine the achievement of social goals within the Sustainable Development Goals (SDGs). A scoping review was used to comprehensively connect AI, social sustainability, and security. Through thematic categorization, the report illustrates how this relationship is framed in research and by official agencies that conceptualize its significance. The study investigates six themes: techno-optimism and AI for social good, economic security, development and humanitarian aid, values, security, and sociotechnical harms. In total, 62 articles and eleven reports are included in the analysis. From this material, six themes were identified: AI for social good, economic security, development and humanitarian aid, values, sociotechnical harms, and societal security.

The article concludes that although various aspects of security are discussed in relation to AI's involvement in achieving the social goals of the SDGs, the concept of security is only partially addressed. When security issues related to AI and social sustainability are discussed, they are often mentioned as indirect effects. The literature that explicitly mentions societal security is primarily concerned with cybersecurity, police work, and financial fraud. However, only one article in this study addresses the issue of national security and the social aspects of technological development in this area. Furthermore, the negative consequences of algorithmic policing and surveillance were not explicitly addressed in the analyzed literature, although these topics could benefit from analysis using a social sustainability perspective. Therefore, one conclusion is that the security context is underdeveloped in this literature, and there is a need for more interdisciplinary research in this area. This article also highlights the importance of value alignment and its social effects as essential for achieving a "good AI society," with a particular emphasis on adopting a human-centric approach. Aligning AI with ethical values is often approached as a regulatory ideal; however, it adds a complexity that is only partially addressed in the analyzed literature.

By combining discussions on AI, security, society, and their interconnected effects on sustainable development, the article highlights the complex nature of synergies and trade-offs that shape policy measures and influence the trajectories of future technological development. Recognizing this connection presents opportunities for developing AI that supports social sustainability while acknowledging the risks associated with emerging technologies that may lead to sociotechnical harms, resulting in heightened polarization, intolerance, and, ultimately, a deteriorated security situation.

This article extends the conceptual framework presented in Article I. It contextualizes AI from a sociotechnical perspective and examines themes in academic work that describe the connection between AI, social sustainability, and security.

4.3 Article III

Ethics as an Enabler and a Constraint – Narratives on Technology Development and Artificial Intelligence in Military Affairs through the Case of Project Maven

Project Maven is an AI-driven information technology initiative for military applications launched by the United States Department of Defense (DoD) in 2017. Initially, it was contracted to a civilian partner, Google. However, this initiative faced significant backlash from many Google employees, leading to the termination of the contract. This article employs narrative analysis to explore the enabling and constraining arguments regarding AI for military purposes that emerged in the debate following the public announcement of Project Maven. Additionally, the article emphasizes the co-production of ethics, technology, and the intricate issues that arise from civilian-military collaborations in technology development. Enabling arguments tied to consequentialist ethics are identified as narratives of accuracy and maintenance. Accuracy serves as a guiding principle for preserving civilian lives, while maintenance aims to maintain the balance of power. In contrast, constraining arguments stem from deontological ethics, highlighting disengagement and ambivalence. Disengagement widens the civilian-military divide, while ambivalence reveals conflicting views on the potential role of technological solutions in warfare. Security narratives and technological storytelling are impactful as they actively shape the framing and mobilization of security and technology development.

This article analyzes and compares socio-technical imaginaries and narratives and how they conflict due to disparate ethical orientations. One particular finding from this study revealed a considerable process of ethicalization of arguments to proceed or halt technology development for dual use. Imaginaries of AI were identified as creating openings and boundaries in discussions about AI applications. This contribution captures value-based tensions in sociotechnical imaginaries, thus illustrating the co-creation of technology, values, and ethical deliberations. The article also describes how this entanglement has contributed to profound ethical dilemmas concerning uncertainty and dual-use technology in today's rapidly evolving sociotechnical landscape.

4.4 Article IV

Death by a Thousand Papercuts - A Qualitative Study on Professionals' Imaginaries of AI, Social Impacts, and Security (Submitted)

Artificial Intelligence, or AI, has driven sociotechnical imaginaries since Turing's era—shaping society's ideal future and revealing how it can stray from that ideal. This study is based on interviews with thirteen respondents from various organizations examining AI's potential to either achieve or hinder societal objectives. By utilizing the sociotechnical imaginaries framework developed by Jasanoff and Kim (2009), along with a narrative approach to AI, this article identifies and evaluates four imaginaries that reflect professionals' perspectives on AI: Reality, Dystopia, Regulation, and Utopia. The key findings reveal a disconnect between current perceptions of the technology and its actual development. Technological change progresses swiftly, yet the visions that guide it remain stuck in a repetitive cycle of binary frameworks, overshadowing pressing issues such as the decline of democratic values, misinformation, and verification challenges. Regulation is another critical theme, complicated by the unpredictability of AI and the limited expertise of policymakers. AI's rapid advancement follows a pattern of innovation driven by capability rather than societal necessity, which can reinforce power imbalances.

A key dimension is the paradox of AI development, which is seen as both an unstoppable, deterministic force and a product of human intent. This duality reflects "sociotechnical blindness," where human choices in technology development are often overlooked. Thus, this article underscores the contradictory perception of AI as an ahistorical technology that operates beyond human influence while simultaneously emphasizing that its capacity for good or bad relies entirely on human intent.

The article also highlights an increasing ethicalization process concerning AI imaginaries, wherein moral arguments are employed to either restrict or promote the advancement of AI. Ethical concerns arise, particularly around values embedded in AI systems, which cause concerns about cultural hegemony and security risks for smaller nations. However, although developers often take ethical implications into account during technology development, they frequently overlook the broader societal implications based on the common assumption that technology is neutral.

This article builds upon the previous articles by examining the conceptual framework through a qualitative empirical study.

5. Discussion

This thesis is guided by the aim of exploring how we can understand social sustainability, security, and AI as a sociotechnical system, along with the implications and perspectives that arise from such an understanding. This aspiration has prompted an examination of the profoundly normative and ontologically varied concepts that interconnect with, as well as co-construct, various norms and values that influence visions of technological progress and the imaginaries that follow. Therefore, a fundamental objective of this thesis has been to discern the impact of social and security aspects on the technology that eventually emerges and to reflect on which imaginaries come to fruition and which remain overlooked. Within this framework, AI demonstrates how sociotechnical imaginaries of security and social sustainability are formed, as well as where their boundaries and openings lie. As discussed in Chapter 2, Theoretical Framework, AI is a ubiquitous general-purpose technology with substantial societal implications. However, it is also a narrativized technology, revealing a significant gap between the *actual* and the *potential.* Additionally, descriptions of AI are often presented in a reductionist manner to invoke emotional responses. In Sherry Turkle's (2004) words:

> It has become a cultural commonplace to use oversimplification about technology as the functional equivalent of political spin – the practice of spinning turns complexity into simple narratives (Turkle, 2004: 19)

As has been argued throughout this thesis, AI is a technology deeply embedded in technological storytelling, presenting conflicting depictions of it as both the solution to social problems and their cause. In other words, the imaginaries of AI have blurred the distinction between predictions about *what could happen* and *what is actually happening*. In the following discussion, I will present three key findings that have been central themes in this thesis: complex sociotechnical systems, ethicalization processes and conflicting values, and, finally, what a co-production perspective adds to the understanding of social sustainability, security, and AI.

Complex Sociotechnical Systems

Complexity lies everywhere in human existence. Yet, when complex phenomena are discussed, they tend to be compartmentalized in neat packages. This is often the case when examining how AI is presented in research (Article II) and societal debates (Article III). However, to fully understand AI's societal impact, the findings

of this thesis suggest that the conversation should proceed by viewing AI and its social effects as a complex sociotechnical system. But what does that mean?

AI, on its own, is a complex and network-based technology. Therefore, even in its most basic form, it encompasses several components: the algorithms created within the AI system, the data used to train the model, and the individuals who interact with it, such as users, designers, and organizational stakeholders. Consequently, a highly complex system is established, which necessitates a reflexive approach to explain the composition of human elements, such as values and biases, combined with purely technological functions in design and applications. More complexity is introduced when adding another normative layer, such as social sustainability and security, which makes analyzing the trajectories of AI, its potential, and its perils a highly intricate matter. Even though an AI for social good initiative is driven by sincere motivation and good intentions, its application may create other problems besides those initially intended to be solved. For instance, from an environmental perspective, AI could help mitigate the global energy crisis (Huang et al., 2020; Levenda et al., 2019), or it could exacerbate the existing energy crisis due to its substantial carbon footprint (Nabavi, Daniell, Williams, and Bentley, 2019). Additionally, numerous queries regarding algorithmic decision-making and just processes arise, which could increase societal security but could also infringe on the goals of social sustainability (Završnik, 2020).

Therefore, evaluating if a desired goal has been aided or impaired by a specific AI application becomes exceedingly complicated. Isolating the goals from one another helps to visualize the connection; however, this operation inhibits a holistic assessment of the interdependent nature of the goals, which is an essential aspect of sustainable development. As discussed in articles II, III, and IV of this thesis, AI systems can be both beneficial and harmful from a societal perspective. Therefore, the outcomes of a sociotechnical system can be challenging to assess beforehand due to a high degree of uncertainty regarding future developments in a vast number of components that change independently yet also alter the system's composition. Furthermore, measuring these components presents another challenge due to the constant input and influence of other variables. However, failing to address complexity can result in unforeseen consequences, including the deepening of structural inequalities in society, which can destabilize societal security.

So, how should we address complex sociotechnical phenomena without compartmentalizing or simplifying their occurrences? This thesis proposes a perspective on imaginaries to approach technology in normative terms, fostering a holistic sociotechnical approach that examines interdependencies and correlations. Although all technology is somewhat sociotechnical, AI possesses a unique self-generating ability, which means it is shaped by societal forces as much as it actively reshapes those forces. At the same time, as technological innovation progresses, AI will eventually be able to reshape itself. Therefore, as a mediator of social norms, AI is highly capable of producing and reinforcing these norms. This specific feature

serves as a fundamental starting point for assessing the capacity of AI as a force for good and advancing the fulfillment of the SDGs while also promoting social goals (Floridi et al., 2018; Saetra, 2022; Theodorou et al., 2022; Vinuesa et al., 2020). However, if not carefully checked, it can steer society in the opposite direction by reinforcing polarization, accelerating structural inequalities, and eroding the fundamental democratic foundations of trust and accountability.

One issue that emerges from acknowledging the composition of social sustainability, security, and AI as a complex sociotechnical system is the assessment and evaluation of potential synergies and trade-offs. This particular aspect is foremostly discussed in Articles I and II. Article I describes the conceptual synergies and trade-offs between social sustainability and security and how they can be approached differently depending on the theoretical perspective applied. In Article II, AI and the social dimension of the SDGs are discussed through a security lens, which emphasizes an interactive systems approach while also highlighting synergies and trade-offs among the goals and targets (Weitz et al., 2018; Yeh et al., 2021).

One finding in this thesis is that much of the evidence of AI impact is derived from experimental closed systems, (Di Vaio et al., 2024; Satornino et al., 2024; Vinuesa et al., 2020), which limits the assessment of the effects of the dynamic relationships between SDG indicators (Saetra, 2021; Weitz et al., 2018). Thus, the possibility of utilizing synergies while evaluating and addressing potential trade-offs is hindered. To assess synergies and trade-offs, this thesis suggests focusing on the essential components of our sociotechnical existence, which takes the formation of values in technological materializations as its starting point.

Therefore, the trade-offs and synergies among specific Sustainable Development Goals (SDGs) can be improved through assessment from a social sustainability perspective, which also considers the security aspect. Although AI-enabled technology can serve as a catalyst for achieving the 2030 Agenda, it may also give rise to unintended consequences, such as increased inequalities, which ultimately impact global and local security.

Ethicalization processes and conflicting values

Because complexity is challenging to address, descriptions of AI's social impact tend to gravitate towards binary descriptions, often formed by various ethicalization processes that are based on techno-utopian or dystopian sentiments. One central theme in this thesis' findings is that the values and ethical considerations associated with AI tend to be at the center of discussions on whether AI has the potential for good or bad. Thus, the political dimension implicit in AI imaginaries is closely tied to ethical constructions that often appear in binary pairs, which can function as both enablers and constraints.

Article III illustrates an ethicalization process through a binary approach to a shared goal of preventing civilian deaths, where AI developed for military use might

increase (or decrease, depending on whom you ask) a nation's security but could lead to a diminution of individual freedom. Because these positions are closely tied to sentiments surrounding existential hopes and fears, they form a solid position that is difficult to alter. Article IV has also identified an ethicalization process surrounding AI imaginaries, where moral arguments are employed to either restrict (through regulation) or promote the advancement of AI. Ethicalization in this context appears as an enabler for sovereign AI, as it raises concerns about cultural hegemony and security risks for smaller nations. Values are described as an influence-weapon embedded in AI systems, with a powerful transformative impact on humans, thus pushing through imaginaries of national LLMs as a necessary counterforce. However, how this influence operates remains unclear, as is whether or not national perspectives on AI are the right cure.

All the articles included in this thesis acknowledge the importance of examining the basis for different value systems, as attuning technology to societal values is not always straightforward. Article I illustrates how the production of values influences the foundation of normative concepts, which in turn impact the basic parameters of the sociotechnical system that follows. Article II emphasizes the importance of value alignment and its social implications as crucial for achieving a "good AI society," with a particular focus on adopting a human-centric approach. However, the concept of "human" is inherently problematic because it contains embedded ethnocentric ideas, and the belief that greater human control over AI systems will reduce bias remains unsettled. Article III examines the effects of two philosophical approaches that create a conflict: accuracy and maintenance as a means to save lives while preserving geopolitical order versus disengagement and ambivalence, which center on existential fears and doubts about AI's potential to preserve humanity at all. Lastly, Article IV shows how developers and professionals in AI are often willing to consider the ethical implications of their work. However, they are not always aware of the complex political and societal problems to which these issues are connected. One explanation for this is that there is a mismatch between the foundational belief that technological development follows a predetermined path, which implies a strong association with technological determinism, and the perspective that human intent—the application of technology—ultimately dictates the parameters for the technology that materializes.

The identification of conflicting norms and values, and the potential effects they may produce, is thus an area that can benefit both the practitioner side and the policy area by enabling a better anticipation of the system's consequences. Aligning AI with ethical values is often approached as a regulatory ideal; however, it introduces a complex dimension where balancing individual preferences with the greater good raises questions about determining whose values should be represented. This dilemma also highlights a classic tension between social goals and AI-based predictions: acknowledging the individual as an autonomous subject while making predictions for the greater good. These questions involve normative assessments of

the values a socially sustainable society should uphold as its ideal and is a central dilemma in sustainability research (Anderson et al., 2016). However, values, too, need to be prioritized, as not all values can be achieved simultaneously. Thus, the process of choosing itself reflects different values (Bromley, 1998).

Adding a co-production perspective

So, what are the benefits of combining a perspective of social sustainability with security when assessing the future potential of AI? Although there has been ample research on obstacles and opportunities related to the SDGs and AI, the connection between security, social sustainability, and AI is not always apparent. This finding is supported by Article II, which concludes that security is usually mentioned as an indirect effect rather than being explicitly mentioned as such. However, this thesis highlights the importance of making this connection when considering the societal implications of AI's potential to achieve social sustainability goals. One example discussed in Articles II and IV is social polarization, which can be linked to various applications of AI systems, both intentional and unintended. Accordingly, the co-production between AI, security, and social sustainability highlights the complexity of issues that need to be addressed to make contributions that will enable society to prosper. Reversing the perspectives, social sustainability can also aid security experts by focusing on maintaining a balance between the human rights of individual citizens and the universal common good, thereby epitomizing questions of security for whom and justice for whom. Nonetheless, as argued in Article I, adding a security perspective requires attuning to the dimensions of politics and power since these dimensions are deeply embedded within the security concept. Failing to do so can reinforce rather than mitigate unequal power positions and prevent broader societal transformation. In this case, thinking through imaginaries can help to detect some of its expressions.

Power and politics are intertwined through various sociotechnical imaginaries that emerge from novel expectations and past experiences, infusing design decisions with the values of particular worldviews (Bowker & Star, 1999). These imaginaries are shaped by underlying values derived from various contexts. However, that diversity tends not to be equally evaluated. Technologies and security measures do not exist independently of an underlying vision. Ultimately, what determines which imaginary prevails is not based solely on technical or engineering factors but is fundamentally about power (Acemoglu and Johnson, 2023). Therefore, the intertwined structures of power, technoscience, social values, and security are essential components to understanding the process by which imaginaries form.

The association between technology and security illustrates a value-based and normative perspective with wide-ranging implications for global security perceptions. Accordingly, technological development and adaptation do not unfold linearly but are formed in a process where values, sociotechnical imaginaries, and technological possibilities co-produce emerging technologies. This composition

opens up opportunities for contributing to a "good society," which can foster ideals associated with social sustainability, such as diversity, social equity, inclusion, and emancipation. However, it also means that these systems have a propensity to breed "sociotechnical harms," where bias can lead to increased polarization, intolerance, and, ultimately, a deteriorated security situation.

6. Conclusions

This thesis has explored how we can comprehend social sustainability, security, and AI as a sociotechnical system and what implications and perspectives arise from such an understanding. This question has been studied at a conceptual level by analyzing how various ontological perspectives can appear simultaneously as harmonious and co-produced, yet also binary and oppositional. The thesis has also examined this aim through various empirical approaches to understand how these conceptualizations manifest in different contexts. A key conclusion from this work, as discussed in the theoretical framework and the appended articles, is that the normative foundation associated with security and social sustainability necessitates a continuous negotiation of what trade-offs are acceptable for a diverse composition of humans and technological elements. Another prominent theme revolves around the interactions between values and technology and how they co-produce the sociotechnical system that takes form. Furthermore, it has been argued throughout this thesis that a sociotechnical perspective, which takes a holistic and integrated viewpoint of various sociological, political, and technological aspects, is a way to expand the conceptual relationship between social sustainability, security, and AI. This perspective thus entails broadening the view from examining specific components in a system to adopting a holistic approach that considers complex systems of synergies and trade-offs focusing on the interaction of ethicalization processes and conflicting values.

6.1 Answering the Research Questions

In this section, I return to the research questions asked in the introduction chapter, starting with RQ 1:

Using a sociotechnical perspective, how can we understand the conceptual relationship between social sustainability, security, and AI?

This question has been investigated in articles I and II. Article I argues that the conceptual relationship between security and social sustainability can be approached from various ontological standpoints, producing different outcomes. Social sustainability and security are typically conceptualized within the UN framework from a co-production perspective. However, this conceptualization has

problems due to the inherent paradoxical elements of the security concept. A deconstructive approach offers an alternative perspective by identifying and challenging binary constructions. Article II examined the relationship between the existing academic literature and AI, social sustainability, and security, concluding that security is typically addressed as an indirect effect rather than a direct approach. Therefore, one conclusion is that, although this relationship is clearly stated from a policy perspective, it is not well-documented in research. Aspects related to security exist, but they need to become more explicit. Conversely, the security domain needs to become more aware of the social elements that impact security. In conclusion, a sociotechnical perspective conceptualizes complex interactions between values, humans, security, and technology. It blurs the boundaries between the social and technological spheres while supporting AI's agential capacity to realize societal values with both positive and negative outcomes.

Based on the theoretical foundation established through RQ1, a second research question emerged, focusing on how this connection is perceived across different contexts. These aspects were pronounced in RQ2:

How is the connection between social sustainability, security, and AI imagined across different contexts?

This research question is investigated in Articles II, III, and IV, which identify imaginaries of social sustainability, security, and AI as profoundly connected with the values and ethicalization processes of AI discourses. Article II has examined the context of academic journals to explore the connection between the social dimension of the SDGs and AI from a security perspective, highlighting its indirect effects. Article III has investigated the debate following the public exposure of Google's involvement in the military Project Maven, which showed how ethics was used to either support or halt the participation of a civilian company in developing AI for military purposes. The case illustrates the growing ethicalization of AI discourses, where ethical arguments are closely tied to the political dimension implicit in AI imaginaries. Ethicalization processes often construct moral principles that appear in binary pairs, which can function as enablers and constraints. In this way, ethical imaginaries play a crucial role in shaping the future trajectories of AI, comprising both contradictory and productive elements.

Article IV investigates the context of AI professionals and identifies imaginaries as constructed through an ethicalization process that highlights the often-overlooked power dynamics embedded in values as arguments to establish national sovereignty in AI development. In addition, this article highlights several value-based inconsistencies; one is the view of AI as merely a tool yet inherently linked to geopolitical issues that can significantly impact the global order. Another tension arises from imagining AI as an ahistorical and neutral technology that follows deterministic patterns, yet simultaneously being a technology deeply rooted in

human intent. This inconsistency arises from the interplay between values and power and is often overshadowed by existential narratives surrounding AI.

In summary, ethical values are essential components in AI imaginaries. Because they are normative and affective, they are challenging to address and alter. However, one way to approach them is to identify their ontological origin and analyze the political motives that shape them.

6.2 Main Contributions

This thesis has contributed by establishing a conceptual foundation for approaching technology development with a holistic perspective on social sustainability and security, offering a more comprehensive view of the opportunities and risks associated with AI. This contribution is aimed at policymaking and serves as a conceptual input for considering complex issues such as AI regulation and its societal impacts.

Additionally, this thesis has advanced the development of conceptualizations in security research by incorporating a social sustainability perspective into the analysis of security-related studies. It has, therefore, contributed to an expanded view of security as closely related to social sustainability.

Another contribution is in the Science, Technology, and Society (STS) field, where this thesis introduces the concept of ethicalization, a critical aspect of the imaginaries surrounding AI. This thesis has demonstrated that ethicalization processes are increasingly influencing both academic and public discussions on the prospects of developing or restricting AI.

At a methodological level, this thesis has contributed to a holistic, issue-oriented approach to addressing research complexity, thereby opening disciplinary boundaries.

6.3 Limitations and Future Research

This thesis has evolved in a bricolage process, meaning the research questions have emerged over time. Therefore, it has changed its course and has not evolved systematically. This is foremostly revealed in the initial application of the sociotechnical framework, where the idea was to study inconsistencies between social sustainability and security, using AI as a case study. However, as the thesis progressed, it became apparent that the system is much more complex than what was initially imagined. Thus, the application became more focused on understanding social sustainability, security, and AI as a holistic system. These two

approaches are presented simultaneously and can, therefore, appear inconsistent in the thesis' theoretical aim. However, since these perspectives coexist in the data on which this thesis is based and also reflect the research process as a whole, I have chosen to let them coexist.

Another limitation is that quantitative and qualitative methods have not been equally balanced, as this thesis has relied chiefly on qualitative methods. That is because the foundation of this thesis is based on a qualitative interpretative tradition. However, that does not mean that the questions asked in this thesis could not have been approached using quantitative methods. It is possible that a quantitative methodology could have been applied to address some of the research questions, potentially revealing another facet of the broader scope of this thesis. A qualitative approach is valuable for allowing a deeper analysis of values and ethical constructions. Still, it is less robust in terms of building generalizability, which is a limitation of this thesis. Another limitation is the critical perspective that has inspired this thesis problem formulation. The critical research tradition is focused on expanding the problem formulation rather than suggesting a way forward. For this thesis, this means that a critical theory perspective has not been fully adopted, and the thesis should not be viewed as strictly following that tradition.

Several research questions have emerged beyond the scope of this thesis during this process. One such question is how imaginaries of ethicalization impact the balancing act between innovation and regulation. The link between economic and social sustainability requires further research, particularly from the perspectives of AI and security, as these can be linked to innovation capacity and societal resilience. Furthermore, since the conceptualization of social sustainability and security is a novel area, it would be productive to refine the theoretical framework by examining specific areas where this holistic approach could be beneficial, such as crisis management, cybersecurity, and crime prevention.

7 References

Acemoglu, D., & Johnson, S. (2023). *Power and progress: Our thousand-year struggle over technology and prosperity*. Basic Books.

Acharya, A. (2001). Human Security: East versus West. *International Journal*, *56*(3), 442–460. https://doi.org/10.1177/002070200105600304

Adams, S. (2023). What are Social Imaginaries?: A Pathway Through the Labyrinth. *International Journal of Social Imaginaries*, *2*(2), 183–212. https://doi.org/10.1163/27727866-bja00031

Adloff, F., & Neckel, S. (2019). Futures of sustainability as modernization, transformation, and control: A conceptual framework. *Sustainability Science*, *14*(4), 1015–1025. https://doi.org/10.1007/s11625-019-00671-2

Ahlqvist, T. (2022). An outline of future-oriented dialectics: Conceptualising dialectical positions, trajectories and processes in the context of futures research. *Futures*, *143*, 103037. https://doi.org/10.1016/j.futures.2022.103037

Al-Homoud, M., & Tassinary, L. G. (2004). Social Interactions at the Neighbourhood-Level as a Function of External Space Enclosure. *Journal of Architectural and Planning Research*, *21*(1), 10–23.

All Nobel Prizes 2024. (n.d.). NobelPrize.Org. Retrieved October 10, 2024, from https://www.nobelprize.org/all-nobel-prizes-2024/

Allen, A. (1999). *The Power of Feminist Theory—Domination, Resistance, Solidarity*. Westview Press.

Amoore, L. (2009). Algorithmic War: Everyday Geographies of the War on Terror. *Antipode*, *41*(1), 49–69. https://doi.org/10.1111/j.1467-8330.2008.00655.x

Amoore, L. (2016). Cloud geographies: Computing, data, sovereignty. *Progress in Human Geography*, *42*(1), 4–24. https://doi.org/10.1177/0309132516662147

Amoore, L. (2020). *Cloud ethics: Algorithms and the attributes of ourselves and others*. Duke University Press.

Amoore, L., & Raley, R. (2016). Securing with algorithms: Knowledge, decision, sovereignty. *Security Dialogue*, *48*(1), 3–10. https://doi.org/10.1177/0967010616680753

Anderson, M. W., Teisl, M. F., & Noblet, C. L. (2016). Whose values count: Is a theory of social choice for sustainability science possible? *Sustainability Science*, *11*(3), 373–383. https://doi.org/10.1007/s11625-015-0345-1

Aradau, C. (2004). Security and the democratic scene: Desecuritization and emancipation. *Journal of International Relations and Development*, *7*(4), 388–413. https://doi.org/10.1057/palgrave.jird.1800030

Aradau, C. (2010). Security That Matters: Critical Infrastructure and Objects of Protection. *Security Dialogue*, *41*(5), 491–514. https://doi.org/10.1177/0967010610382687

Aradau, C., Coward, M., Herschinger, E., Owen, T., & Voelkner, N. (2015). Discourse/Materiality. In C. Aradau, J. Huysmans, A. Neal, & N. Voelkner (Eds.), *Critical Security Methods: New Frameworks for Analysis* (pp. 57–84). Routledge/Taylor and Francis.

Aradau, C., Huysmans, J., Neal, A., & Voelkner, N. (2014). Introducing critical security methods. In *Critical Security Methods* (pp. 1–22). Routledge.

Arendt, H. (1958). *The human condition*. Chicago, Ill

Åsberg, C., & Braidotti, R. (2018). *A Feminist Companion to the Posthumanities*. Springer.

Åsberg, C., & Lykke, N. (2010). Feminist technoscience studies. *European Journal of Women's Studies*, *17*(4), 299–305. https://doi.org/10.1177/1350506810377692

Bächle, T. C., & Bareis, J. (2022). "Autonomous weapons" as a geopolitical signifier in a national power play: Analysing AI imaginaries in Chinese and US military policies. *European Journal of Futures Research*, *10*(1), 20. https://doi.org/10.1186/s40309-022-00202-w

Backman, S. (2023). Risk vs. Threat-based cybersecurity: The case of the EU. *European Security*, *32*(1), 85–103. https://doi.org/10.1080/09662839.2022.2069464

Bal, M., & Boheemen, C. van. (2017). *Narratology: Introduction to the theory of narrative*. University Of Toronto Press.

Barad, K. (2007). *Meeting the Universe Halfway Quantum Physics and the Entanglement of Matter and Meaning*. Duke University Press.

Beck, S., Jasanoff, S., Stirling, A., & Polzin, C. (2021). The governance of sociotechnical transformations to sustainability. *Current Opinion in Environmental Sustainability*, *49*, 143–152. https://doi.org/10.1016/j.cosust.2021.04.010

Beer, D. (2009). Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media & Society*, *11*(6), 985–1002. https://doi.org/10.1177/1461444809336551

Beer, D. (2017). The social power of algorithms. *Information, Communication & Society*, *20*(1), 1–13. https://doi.org/10.1080/1369118X.2016.1216147

Bengio, Y. (2023). AI and Catastrophic Risk. *Journal of Democracy*, *34*(4), 111–121. https://doi.org/10.1353/jod.2023.a907692.

Bennett, J. (2010). *Vibrant Matter. A Political Ecology of Things*. Duke University Press.

Birhane, A., Kalluri, P., Card, D., Agnew, W., Dotan, R., & Bao, M. (2022). The Values Encoded in Machine Learning Research. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 173–184. https://doi.org/10.1145/3531146.3533083

Bonditti, P., Neal, A., Opitz, S., & Zebrowski, C. (2014). Genealogy. In *Critical Security Methods*. 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. https://doi.org/10.1080/09537320600777002

Boström, N. (2024). *Deep Utopia—Life and meaning in a solved world*. Ideapress Pub.

Bourbeau, P. (2015). Resilience and International Politics: Premises, Debates, Agenda. *International Studies Review*, *17*(3), 374–395. https://doi.org/10.1111/misr.12226

Bousquet, A. (2022). *The scientific way of warfare: Order and chaos on the battlefields of modernity*. Oxford Univ. Press.

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. MIT Press.

Brenner, B., & Hartl, B. (2021). The perceived relationship between digitalization and ecological, economic, and social sustainability. *Journal of Cleaner Production*, *315*, 128128. https://doi.org/10.1016/j.jclepro.2021.128128

Brey, P. (2018). The strategic role of technology in a good society. *Technology in Society*, *52*, 39–45. https://doi.org/10.1016/j.techsoc.2017.02.002

Bromley, D. W. (1998). Searching for sustainability: The poverty of spontaneous order. *Ecological Economics*, *24*(2), 231–240. https://doi.org/10.1016/S0921-8009(97)00145-6

Browne, S. (2015). *Dark matters: On the surveillance of blackness*. Duke University Press.

Brundtland, G. H. (1987). *Our common future*. [United Nations Environment Programme]. Oxford University Press

Bryman, A. (2015). *Social Research Methods*. Oxford University Press.

Bucher, T. (2018). *If...then: Algorithmic power and politics*. Oxford University Press.

Bueger, C., & Stockbruegger, J. (2018). Actor-Network Theory. In M. D (Ed.), *Technology and World Politics* (pp. 60–84). Routledge, Taylor & Francis Group.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 2053951715622512. https://doi.org/10.1177/2053951715622512

Burton, J. (2023). Algorithmic extremism? The securitization of artificial intelligence (AI) and its impact on radicalism, polarization, and political violence. *Technology in Society*, 102262. https://doi.org/10.1016/j.techsoc.2023.102262

Butler, J. (1990). *Gender Trouble: Feminism and the Subversion of Gender*. Routledge/Taylor and Francis.

Butler, J. (2009). Performativity, Precarity and Sexual Politics. *AIBR Revista de Antropologia Ibo Americana* (3).

Buzan, B., Wæver, O., & Wilde, J. de. (1998). *Security: A new framework for analysis*. Lynne Rienner.

Byrne, D. (1998). *Complexity Theory and the Social Sciences: An Introduction*. Routledge. https://search.ebscohost.com/login.aspx?direct=true&db=e000xww&AN=60861&site=ehost-live

Callon, M. (2009). *Acting in an uncertain world: An essay on technical democracy* (P. Lascoumes & Y. Barthe, Eds.). Cambridge, Mass. : MIT Press.

Cave, S., Dihal, K., & Dillon, S. (2020). *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*. Oxford University Press.

Chandler, D. (2008). Human Security II: Waiting for the Tail To Wag the Dog—A Rejoinder to Ambrosetti, Owen and Wibben. *Security Dialogue*, *39*(4), 463–469. https://doi.org/10.1177/0967010608094041

Cheney-Lippold, J. (2017). *We are data: Algorithms and the making of our digital selves*. New York University Press.

Chong, D. (2000). *Rational Lives: Norms and Values in Politics and Soceity*. The University of Chicago Press.

CHS. (2003). *Human Security Now*. Commission on Human Security.

Clarke, V., & Braun, V. (2017). Thematic analysis. *The Journal of Positive Psychology*, *12*(3), 297–298. https://doi.org/10.1080/17439760.2016.1262613

Cockburn, C. (1985). *Machinery of Dominance: Women, Men and Technical Know-How*. Pluto Press.

Coeckelbergh, M. (2023). Democracy, epistemic agency, and AI: political epistemology in times of artificial intelligence. *AI and Ethics*, *3*(4), 1341–1350. https://doi.org/10.1007/s43681-022-00239-4

Coeckelbergh, M., & Gunkel, D. J. (2023). ChatGPT: deconstructing the debate and moving it forward. *AI & SOCIETY*. https://doi.org/10.1007/s00146-023-01710-4

Collier, D., Daniel Hidalgo, F., & Olivia Maciuceanu, A. (2006). Essentially contested concepts: Debates and applications. *Journal of Political Ideologies*, *11*(3), 211–246. https://doi.org/10.1080/13569310600923782

Collingridge, D. (1980). *The Social Control of Technology*. St. Martin's Press.

Collins, H. (2021). The science of artificial intelligence and its critics. *Interdisciplinary Science Reviews*, *46*(1–2), 53–70. https://doi.org/10.1080/03080188.2020.1840821

Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press.

Crabtree, A. (2020). *Sustainability, capabilities and human security*. Springer International Publishing.

Crawford, K. (2021). *Atlas of AI : power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Crutzen, P. J. (2006). *The "Anthropocene" BT - Earth System Science in the Anthropocene* (E. Ehlers & T. Krafft, Eds.; pp. 13–18). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-26590-2_3

Dauphinee, E. (2013). *The politics of exile*. Routledge.

Delanty, G. (2021). Futures of sustainability: Perspectives on social imaginaries and social transformation. A comment on Frank Adloff and Sighard Neckel's research program. *Social Science Information*, *60*(2), 285–294. https://doi.org/10.1177/0539018421999562

Dempsey, N., Bramley, G., Power, S., & Brown, C. (2011). The social dimension of sustainable development: Defining urban social sustainability. *Sustainable Development*, *19*(5), 289–300. https://doi.org/doi:10.1002/sd.417

Desmet, M. (2022). *The psychology of totalitarianism*. Chelsea Green Publishing.

Deveaux, M. (1994). Feminism and Empowerment: A Critical Reading of Foucault. *Feminist Studies*, *20*(2), 223–247. https://doi.org/10.2307/3178151

DeWitt, R. (2018). *Worldviews: An introduction to the history and philosophy of science*. Wiley.

Dhrago, M. (2024). *The Future of European Competitiveness* (pp. 1–69). European Commission. https://commission.europa.eu/topics/strengthening-european-competitiveness/eu-competitiveness-looking-ahead_en

Di Vaio, A., Zaffar, A., & Balsalobre-Lorente, D. (2024). Carbon and Decarbonization Disclosure: Role of Responsible Innovation in Adoption of Artificial Intelligence of Things Towards SDGs. In S. Misra, K. Siakas, & G. Lampropoulos (Eds.), *Artificial Intelligence of Things for Achieving Sustainable Development Goals* (pp. 99–121). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-53433-1_6

Douglas, H. E. (2009). *Science, policy, and the value-free ideal*. University of Pittsburgh Press.

Duffield, M. R. (2007). *Development, security and unending war: Governing the world of peoples*. Polity.

Edwards, M. G. (2021). The Growth Paradox, Sustainable Development, and Business Strategy *Business Strategy and the Environment*, *30*(7), 3079–3094. https://doi.org/10.1002/bse.2790

Egbert, S., & Leese, M. (2021). *Criminal futures: Predictive policing and everyday police work*. Routledge.

Elkington, J. (2008). The Triple Bottom Line. In M. V. Russo (Ed.), *Enviromental Management—Readings and Cases* (pp. 49–68). SAGE Publications.

Elliott, K. C. (2017). *A tapestry of values: An introduction to values in science*. Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780190260804.001.0001

Eubanks, V. (2017). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Faulkner, W. (2001). The technology question in feminism: A view from feminist technology studies. *Women's Studies International Forum*, *24*(1), 79–95. https://doi.org/10.1016/S0277-5395(00)00166-7

Feenberg, A. (2002). *Transforming Technology—A Critical Theory Revisited*. Oxford University Press.

Feenberg, A. (2017). A Critical Theory of Technology. In U. Felt, R. Fouché, C. M. Miller, & L. Smith-Doerr (Eds.), *Handbook of Science and Technology Studies.* (pp. 635–663). MIT Press.

Feenberg, A., & Beira, E. (2018). *Technology, modernity, and democracy: Essays by Andrew Feenberg*. Rowman & Littlefield International.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Focault, M. (1997). *Soceity Must be Defended Lectures at the Collége de France 1975-76*. Picador.

Ford, M. C. (2017). *Weapon of choice: Small arms and the culture of military innovation*. Hurst & Company.

Fosso Wamba, S., Bawack, R. E., Guthrie, C., Queiroz, M. M., & Carillo, K. D. A. (2021). Are we preparing for a good AI society? A bibliometric review and research agenda. *Technological Forecasting and Social Change*, *164*, 120482. https://doi.org/10.1016/j.techfore.2020.120482

Friedman, B. (1997). *Human Values and the Design of Computer Technology* (B. Friedman, Ed.). Cambridge University Press.

Fujii, L. A. (2010). Shades of truth and lies: Interpreting testimonies of war and violence. *Journal of Peace Research*, *47*(2), 231–241. https://doi.org/10.1177/0022343309353097

Gabriel, I. (2020). Artificial Intelligence, Values, and Alignment. *Minds and Machines*, *30*(3), 411–437. https://doi.org/10.1007/s11023-020-09539-2

Gawell, M. (2013). Social Entrepreneurship: Action Grounded in Needs, Opportunities and/or Perceived Necessities? *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, *24*(4), 1071–1090. https://doi.org/10.1007/s11266-012-9301-1

Geels, F. W. (2004). From sectoral systems of innovation to socio-technical systems. Insights about dynamics and change from sociology and institutional theory. *Research Policy*, *33*, 897–920. https://doi.org/10.1016/j.respol.2004.01.015

Geels, F. W. (2005). *Technological transitions and system innovations: A co-evolutionary and socio-technical analysis*. Edward Elgar.

Gehl Sampath, P. (2021). Governing Artificial Intelligence in an Age of Inequality. *Global Policy*, *12*(S6), 21–31. https://doi.org/10.1111/1758-5899.12940

Gerhold, L., & Brandes, E. (2021). Sociotechnical imaginaries of a secure future. *European Journal of Futures Research*, *9*(1), 7. https://doi.org/10.1186/s40309-021-00176-1

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. Boczkowski, & K. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167–195). MIT Press.

Goh, H. H., & Vinuesa, R. (2021). Regulating artificial-intelligence applications to achieve the sustainable development goals. *DISCOVER SUSTAINABILITY*, *2*(1). https://doi.org/10.1007/s43621-021-00064-5

Goralski, M. A., & Tan, T. K. (2022). Artificial intelligence and poverty alleviation: Emerging innovations and their implications for management education and sustainable development. *INTERNATIONAL JOURNAL OF MANAGEMENT EDUCATION*, *20*(3). https://doi.org/10.1016/j.ijme.2022.100662

Graham, M. (2020). Materiality. *Lambda Nordica*, *25*(1 SE-Articles), 87–91. https://doi.org/10.34041/ln.v25.618

Gunkel, D. (2012). *The Machine Question Critical Perspectives on AI, Robots, and Ethics*. MIT Press.

Guzzini, S. (2000). A Reconstruction of Constructivism in International Relations. *European Journal of International Relations*, *6*(2), 147–182. https://doi.org/10.1177/1354066100006002001

Haddon, L. (2000). Social Exclusion and Information and Communication Technologies: Lessons from Studies of Single Parents and the Young Elderly. *New Media & Society*, *2*(4), 387–406. https://doi.org/10.1177/1461444800002004001

Hagbert, P., Wangel, J., & Broms, L. (2020). Exploring the Potential for Just Urban Transformations in Light of Eco-Modernist Imaginaries of Sustainability. *Urban Planning*, *5*(4), 204–216. https://doi.org/10.17645/up.v5i4.3302

Hall, R. (2015). *The Transparent Traveller: The Performance and Culture of Airport Security*. Duke University Press.

Hanlon, R. J., & Christie, K. (2016). Human Security, Conflict, and Development. In *Freedom from Fear, Freedom from Want* (pp. 3–22). University of Toronto Press.

Hansson, S. O. (2015). Science and Technology: What They Are and Why Their Relation Matters. In S. O. Hansson (Ed.), *The Role of Technology in Science: Philosophical Perspectives. Philosophy of Engineering and Technology, vol 18*. Springer. https://doi.org/10.1007/978-94-017-9762-7_2

Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, *14*(3), 575–599. https://doi.org/10.2307/3178066

Haraway, D. (1991). *Simians, Cyborgs, and Women: The Reinvention of Nature*. Free Association Books.

Harding, S. (1986). *The Science Question in Feminism* (C. U. Press, Ed.).

Harrag, F., & Alshehri, A. (2023). Applying Data Mining in Surveillance: Detecting Suspicious Activity on Social Networks. *International Journal of Distributed Systems and Technologies (IJDST)*, *14*(1), 1–24.

Hartman, J., Butts, K., Bankus, B., & Carney, S. (2012). *Sustainability and National Security*. Center for Strategic Leadership United States Army War College.

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-Inspired Artificial Intelligence. *Neuron*, *95*(2), 245–258. https://doi.org/10.1016/j.neuron.2017.06.011

Hassel, H., Cedergren, A., & Tehler, H. (2022). Strategies for Aggregating Risk Information in a Societal Safety Context. *Safety Science*, *149*, 105691. https://doi.org/10.1016/j.ssci.2022.105691

Hayles, K. (1999). *How We Became Posthuman*. The University of Chicago Press.

Hitlin, S., & Piliavin, J. A. (2004). Values: Reviving a Dormant Concept. *Annual Review of Sociology*, *30*, 359–393. https://doi.org/10.1146/annurev.soc.30.012703.110640

Holton, J., & Walsh, I. (2017). *Classic grounded theory: Applications with qualitative and quantitative data*. Sage.

Hoogensen Gjorv, G. (2012). Security by any other name: Negative security, positive security, and a multi-actor security approach. *Review of International Studies*, *38*(4), 835–859. https://doi.org/DOI: 10.1017/S0260210511000751

Horowitz, M. (2018). Artificial Intelligence, International Competition, and the Balance of Power. *Texas National Security Review*, *1*(3), 38–57. https://doi.org/10.15781/T2639KP49

Houkes, W. (2009). The Nature of Technological Knowledge. In A. Meijers (Ed.), *Handbook of the Philosophy of Science* (pp. 309–350). North-Holland. https://doi.org/10.1016/B978-0-444-51667-1.50016-1

Høyland, S. A. (2018). Exploring and modeling the societal safety and societal security concepts – A systematic review, empirical study and key implications. *Safety Science*, *110*, 7–22. https://doi.org/10.1016/j.ssci.2017.10.019

Hummel, P., Braun, M., & Dabrock, P. (2021). Own Data? Ethical Reflections on Data Ownership. *Philosophy & Technology*, *34*(3), 545–572. https://doi.org/10.1007/s13347-020-00404-9

Huysmans, J. (1998). Security! What Do You Mean?: From Concept to Thick Signifier. *European Journal of International Relations*, *4*(2), 226–255. https://doi.org/10.1177/1354066198004002004

Jackson, P. (2016). *The Conduct of Inquiry in International Relations: Philosophy of Science and its Implications for the Study of World Politics*. Routledge/Taylor and Francis.

James, W. (1909). *A pluralistic universe: Hibbert lectures at Manchester College on the present situation in philosophy*. Longmans, Green.

Jasanoff, S. (2015). Future Imperfect: Science, Technology and the Imaginations of Modernity. In S. Jasanoff & S.-H. Kim (Eds.), *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power* (pp. 1–33). University of Chicago Press.

Jasanoff, S. (2016). *The ethics of invention: Technology and the human future*. W.W. Norton & Company.

Jasanoff, S., & Kim, S.-H. (2009). Containing the Atom: Sociotechnical Imaginaries and Nuclear Power in the United States and South Korea. *Minerva*, *47*(2), 119–146. https://doi.org/10.1007/s11024-009-9124-4

Johnson, J. (2019). Artificial intelligence & future warfare: Implications for international security. *Defense & Security Analysis*, *35*(2), 147–169. https://doi.org/10.1080/14751798.2019.1600800

Joly, P.-B. (2015). Governing Emerging Technologies; The need to think out of the (black) box. In S. Hilgartner, C. Miller, & R. Hagendjik (Eds.), *Science and Democracy—Making Knowledge and Making Power in the Biosciences and Beyond* (pp. 133–156). Routledge.

Juelskjær, M., & Schwennesen, N. (2012). Intra-active Entanglements – An Interview with Karen Barad. *Kvinder, Køn & Forskning*, *1–2*. https://doi.org/10.7146/kkf.v0i1-2.28068

Kalantzakos, S. (2020). The Race for Critical Minerals in an Era of Geopolitical Realignments. *The International Spectator*, *55*(3), 1–16. https://doi.org/10.1080/03932729.2020.1786926

Katrina Ligett, Terah Lyons, James Manyika, Juan Carlos Niebles, Yoav Shoham, Russell Wald, and J. C. (2024). *The AI Index 2024 Annual Report*. AI Index Steering Committee, Institute for Human-Centered AI.

Kenter, J. O., Raymond, C. M., van Riper, C. J., Azzopardi, E., Brear, M. R., Calcagni, F., Christie, I., Christie, M., Fordham, A., Gould, R. K., Ives, C. D., Hejnowicz, A. P., Gunton, R., Horcea-Milcu, A.-I., Kendal, D., Kronenberg, J., Massenberg, J. R., O'Connor, S., Ravenscroft, N., … Thankappan, S. (2019). Loving the mess: Navigating diversity and conflict in social values for sustainability. *Sustainability Science*, *14*(5), 1439–1461. https://doi.org/10.1007/s11625-019-00726-4

King, G., & Murray, C. J. L. (2001). Rethinking Human Security. *Political Science Quarterly*, *116*(4), 585–610. https://doi.org/10.2307/798222

Kissinger, H. A. (1961). *The Necessity for Choice*. Harper.

Konrad, K., & Böhle, K. (2019). Socio-technical futures and the governance of innovation processes—An introduction to the special issue. *Futures*, *109*, 101–107. https://doi.org/10.1016/j.futures.2019.03.003

Korsgaard, C. (1996). *The Sources of Normativity*. Cambridge University Press.

Kroes, P. (2012). Technical Artefacts: Creations of Mind and Matter: A Philosophy of Engineering Design. In *Philosophy of Engineering and Technology 6* (pp. 13–45). Springer. https://doi.org/10.1007/978-94-007-3940-6_2

Kudina, O., & Verbeek, P.-P. (2018). Ethics from Within: Google Glass, the Collingridge Dilemma, and the Mediated Value of Privacy. *Science, Technology, & Human Values*, *44*(2), 291–314. https://doi.org/10.1177/0162243918793711

Kurzweil, R. (2004). *The Law of Accelerating Returns BT - Alan Turing: Life and Legacy of a Great Thinker* (C. Teuscher, Ed.; pp. 381–416). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-05642-4_16

Latour, B. (1992). Where Are the Missing Masses? In W. Bijker & J. Law (Eds.), *Shaping Technology/Building Society: Studies in Sociotechnical Change* (pp. 225–258). Cambridge University Press.

Latour, B. (2004). *Politics of Nature*. Harvard University Press.

Law, J. (1990). Introduction: Monsters, Machines and Sociotechnical Relations. *The Sociological Review*, *38*(1_suppl), 1–23. https://doi.org/10.1111/j.1467-954X.1990.tb03346.x

Law, J. (2004). *After Method: Mess in Social Science*. Routledge.

Law, J., & Singleton, V. (2000). Performing Technology's Stories: On Social Constructivism, Performance, and Performativity. *Technology and Culture*, *41*(4), 765–775.

Law, J., & Singleton, V. (2005). Object Lessons. *Organization*, *12*(3), 331–355. https://doi.org/10.1177/1350508405051270

Lawrence, R. J. (2010). Deciphering Interdisciplinary and Transdisciplinary Contributions. *Transdisciplinary Journal of Engineering & Science*, *1*(0 SE-Articles). https://doi.org/10.22545/2010/0003

Lawson, S. (2011). Articulation, antagonism, and intercalation in Western military imaginaries. *Security Dialogue*, *42*(1), 39–56. https://doi.org/10.1177/0967010610393775

Leavy, P. (2011). *Essentials of transdisciplinary research: Using problem-centered methodologies*. Left Coast Press.

Leese, M. (2017). Holding the Project Accountable: Research Governance, Ethics, and Democracy. *Science and Engineering Ethics*, *23*(6), 1597–1616. https://doi.org/10.1007/s11948-016-9866-y

Leese, M. (2019). Configuring Warfare—Automation, Control, Agency. In M. Hoijtink & M. Leese (Eds.), *Technology and agency in international relations* (pp. 42–66). Routledge.

Leins, K., Arendt, H., & Weizenbaum, J. (2019). AI for better or for worse, or AI at all. In *Artificial Intelligence for Better or Worse* (pp. 3–17). Future Leaders.

Littig, B. (2002). The Case for Gender-sensitive Socio-ecological Research. *Work, Employment and Society*, *16*(1), 111–132. https://doi.org/10.1177/09500170222119272

Litting, B., & Griesler, E. (2005). Social sustainability: A catchword between political pragmatism and social theory. *International Journal for Sustainable Development*, *8*(1/2), 65–79. https://doi.org/10.1504/IJSD.2005.007375

Liwång, H., Andersson, K. E., Bang, M., Malmio, I., & Tärnholm, T. (2023). How can systemic perspectives on defence capability development be strengthened? *Defence Studies*, *23*(3), 399–420. https://doi.org/10.1080/14702436.2023.2239722

Loh, J. (2019). What is feminist philosophy of technology? A critical overview and a plea for a feminist technoscientific utopia. In J. Loh & M. Coeckelbergh (Eds.), *Feminist philosophy of technology*. Springer Berlin / Heidelberg.

Los, M. (2004). The technologies of Total Domination. *Surveillance & Society*, *2*(1), 15–38. https://doi.org/10.24908/ss.v2i1.3325

Lustig, C., Pine, K., Nardi, B., Irani, L., Lee, M. K., Nafus, D., & Sandvig, C. (2016). Algorithmic Authority: The Ethics, Politics, and Economics of Algorithms That Interpret, Decide, and Manage. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 1057–1062. https://doi.org/10.1145/2851581.2886426

MacKenzie, D. (1993). *Inventing Accuracy—A Historical Sociology of Nuclear Missile Guidance*. MIT Press.

MacKenzie, D., & Wajcman, J. (1999). *The Social Shaping of Technology* (D. MacKenzie & J. Wajcman, Eds.). Open University Press.

Macnish, K. (2018). *The Ethics of Surveillance*. Routledge.

Magis, K. (2010). Community Resilience: An Indicator of Social Sustainability. *Society & Natural Resources*, *23*(5), 401–416. https://doi.org/10.1080/08941920903305674

Malmio, I., & Liwång, H. (2022). Education for Sustainable Development. In W. Leal Filho, A. M. Azul, F. Doni, & A. L. Salvia (Eds.), *BT - Handbook of Sustainability Science in the Future: Policies, Technologies and Education by 2050* (pp. 1–21). Springer International Publishing. https://doi.org/10.1007/978-3-030-68074-9_126-1

Manjikian, M. (2018). Social Construction of Technology- How objects acquire meaning in society. In D. McCharty (Ed.), *Technology and World Politics* (pp. 25–41). Routledge, Taylor & Francis Group.

Marcuse, H. (1968). *Negotiations: Essays in Critical Theory*. Allen Lane.

Martins, B. O., & Ahmad, N. (2020). The security politics of innovation: Dual-use technology in the EU's security research programme. In *Emerging Security Technologies and EU Governance*. Routledge.

Mau, S., & Howe, S. (2019). *The metric society: On the quantification of the social*. Polity Press.

McCarthy, D. R. (2017). Introduction: Technology in world politics. In *Technology and World Politics*. Routledge.

McKenzie, S. (2004). *Social Sustainability: Towards Some Definitions*. Hawke Research Institute.

Meadows, D. (1997). Places to Intervene in a System. *Whole Earth*, *91*, 78–84.

Mecklin, J. (n.d.). 2025 Doomsday Clock Statement. *Bulletin of the Atomic Scientists*. Retrieved February 10, 2025, from https://thebulletin.org/doomsday-clock/2025-statement/

Mensah, J. (2019). Sustainable development: Meaning, history, principles, pillars, and implications for human action: Literature review. *Cogent Social Sciences*, *5*(1), 1653531. https://doi.org/10.1080/23311886.2019.1653531

Meredith, J. (1993). Theory Building through Conceptual Methods. *International Journal of Operations & Production Management*, *13*(5), 3–11. https://doi.org/10.1108/01443579310028120

Merleau-Ponty, M., & Landers, D. A. (2014). *Phenomenology of Perception*. Routledge.

Meyer, U. (2019). The emergence of an envisioned future. Sensemaking in the case of "Industrie 4.0" in Germany. *Futures*, *109*, 130–141. https://doi.org/10.1016/j.futures.2019.03.001

Mhurchú, A. N., & Shindo, R. (2016). *Critical Imaginations in International Relations*. Routledge.

Michelfelder, P. D., Wellner, G., & Wiltse, H. (2017). Designing Differently: Toward a Methodology for an Ethics of Feminist Technology Design. In S. O. Hansson (Ed.), *The Ethics of Technology—Methods and Approaches* (pp. 193–218). Rowman & Littlefield.

Miller, T. R. (2020). Imaginaries of Sustainability: The Techno-Politics of Smart Cities. *Science as Culture*, *29*(3), 365–387. https://doi.org/10.1080/09505431.2019.1705273

Mitcham, C. (1994). *Thinking through technology- The path between engineering and philosophy*. The University of Chicago Press.

Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology*, *33*(4), 659–684. https://doi.org/10.1007/s13347-020-00405-8

Moon, M. J. (2023). Searching for inclusive artificial intelligence for social good: Participatory governance and policy recommendations for making AI more inclusive and benign for society. *Public Administration Review*. https://doi.org/10.1111/puar.13648

Morgenthau, H. J., & Thompson, K. W. (1993). *Politics among nations: The struggle for power and peace*. McGraw-Hill.

Moroni, S. (2019). The just city. Three background issues: Institutional justice and spatial justice, social justice and distributive justice, concept of justice and conceptions of justice. *Planning Theory*, *19*(3), 251–267. https://doi.org/10.1177/1473095219877670

Mügge, D. (2024). EU AI sovereignty: For whom, to what end, and to whose benefit? *Journal of European Public Policy*, *31*(8), 2200–2225. https://doi.org/10.1080/13501763.2024.2318475

Müller, F. I., & Richmond, M. A. (2023). The technopolitics of security: Agency, temporality, sovereignty. *Security Dialogue*, *54*(1), 3–20. https://doi.org/10.1177/09670106221141373

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, *18*(1), 143. https://doi.org/10.1186/s12874-018-0611-x

Murakami, H., & Gabriel, P. (2005). *Kafka on the shore*. Harvill Press.

Navas, E., Gallagher, O., & Burrough, X. (2018). *Keywords in remix studies*. Routledge.

Nayak, A., & Chia, R. (2011). Thinking becoming and emergence: Process philosophy and organization studies. In H. Tsoukas & R. Chia (Eds.), *Philosophy and Organization Theory* (Vol. 32, pp. 281–309). Emerald Group Publishing Limited. https://doi.org/10.1108/S0733-558X(2011)0000032012

Nelson, D. M. (2024). Hydroelectric Chimeras and "Our" Mayan Rivers: De-inscribing Security in Guatemala. *Science, Technology, & Human Values*, 01622439231225531. https://doi.org/10.1177/01622439231225531

Nye, J. S. (2004). *Soft power: The means to success in world politics*. Public Affairs.

Öberg, D. (2019). Ethics, the Military Imaginary, and Practices of War. *Critical Studies on Security*, *7*(3), 199–209. https://doi.org/10.1080/21624887.2019.1672482

OECD. (2023). *Recommendation of the Council on Artificial Intelligence*. OECD/LEGAL/0449. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

Olsen, O. E., Kruke, B., & Hovden, J. (2007). Societal Safety: Concept, Borders and Dilemmas. *Journal of Contingencies and Crisis Management*, *15*. https://doi.org/10.1111/j.1468-5973.2007.00509.x

Orike, S., & Ene, D. S. (2023). Meta-Learning: Unleashing the Power of Self-Improving Artificial Intelligent (AI) Systems. *Journal of Advances in Computational Intelligence Theory*, *5*(3), 12–27. https://doi.org/DOI: https://doi.org/10.5281/zenodo.8223579

Osoba, O., Boudreaux, B., Saunders, J., Irwin, L. J., Mueller, P. A., & Cherney, S. (2019). *Algorithmic Equity. A Framework for Social Applications*. RAND Corporation.

Padden, M., & Öjehag-Pettersson, A. (2021). Protected how? Problem representations of risk in the General Data Protection Regulation (GDPR). *Critical Policy Studies*, *15*(4), 486–503. https://doi.org/10.1080/19460171.2021.1927776

Paris, R. (2001). Human Security: Paradigm Shift or Hot Air? *International Security*, *26*(2), 87–102.

Park-Kang, S. (2015). Fictional IR and imagination: Advancing narrative approaches. *Review of International Studies*, *41*(2), 361–381. https://doi.org/DOI: 10.1017/S0260210514000291

Pasquale, F. (2015). *The Black Box Society*. Harvard University Press.

Peoples, C. (2011). Security after emancipation? Critical Theory, violence and resistance. *Review of International Studies*, *37*(3), 1113–1135.

Peoples, C., & Vaughan-Williams, N. (2020). *Critical Security Studies*. Routledge.

Peters, M. D. J., Marnie, C., Tricco, A. C., Pollock, D., Munn, Z., Alexander, L., McInerney, P., Godfrey, C. M., & Khalil, H. (2020). Updated methodological guidance for the conduct of scoping reviews. *JBI Evidence Synthesis*, *18*(10). https://journals.lww.com/jbisrir/fulltext/2020/10000/updated_methodological_guidance_for_the_conduct_of.4.aspx

Pinch, T. J., & Bijker, W. E. (1984). The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology might Benefit Each Other. *Social Studies of Science*, *14*(3), 399–441. https://doi.org/10.1177/030631284014003004

Polkinghorne, D. E. (2000). Psychological Inquiry and the Pragmatic and Hermeneutic Traditions. *Theory & Psychology*, *10*(4), 453–479. https://doi.org/10.1177/0959354300104002

Programme), U. (United N. D. (1994). Human Development Report 1994. *UNDP (United Nations Development Programme)*.

Pulido, L., & Peña, D. (1998). Environmentalism and Positionality: The Early Pesticide Campaign of the United Farm Workers' Organizing Committee, 1965-71. *Race, Gender & Class*, *6*(1), 33–50.

Radder, H. (2019). *From Commodification to the Common Good: Reconstructing Science, Technology, and Society*. University of Pittsburgh Press.

Ramazanoglu, C., & Holland, J. (2002). *Feminist Methodology: Challenges and Choices*. 1–196.

Ravecca, P., & Dauphinee, E. (2018). Narrative and the Possibilities for Scholarship. *International Political Sociology*, *12*(2), 125–138. https://doi.org/10.1093/ips/olx029

Redclift, M. (2005). Sustainable development (1987–2005): An oxymoron comes of age. *Sustainable Development*, *13*(4), 212–227. https://doi.org/10.1002/sd.281

Rosa, H., & Trejo-Mathys, J. (2013). *Social acceleration: A new theory of modernity*. Columbia University Press.

Rose, J., & Cachelin, A. (2018). Critical sustainability: Incorporating critical theories into contested sustainabilities. *Journal of Environmental Studies and Sciences*, *8*(4), 518–525. https://doi.org/10.1007/s13412-018-0502-9

Roßmann, M. (2021). Vision as make-believe: How narratives and models represent sociotechnical futures. *Journal of Responsible Innovation*, *8*(1), 70–93. https://doi.org/10.1080/23299460.2020.1853395

Rothschild, E. (1995). What Is Security? *Daedalus*, *124*(3), 53–98.

Runciman, D. (2019). *How democracy ends*. Profile Books.

Ryan, M. (2024). We're only human after all: A critique of human-centred AI. *AI & SOCIETY*. https://doi.org/10.1007/s00146-024-01976-2

Rychnovská, D. (2016). Governing dual-use knowledge: From the politics of responsible science to the ethicalization of security. *Security Dialogue*, *47*(4), 310–328. https://doi.org/10.1177/0967010616658848

Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, *6*(1), 2053951718820549. https://doi.org/10.1177/2053951718820549

Saetra, H. S. (2021). AI in Context and the Sustainable Development Goals: Factoring in the Unsustainability of the Sociotechnical System. *SUSTAINABILITY*, *13*(4). https://doi.org/10.3390/su13041738

Saetra, H. S. (2022). *AI for the Sustainable Development Goals*. Bristol University Press.

Sahlgren, M., & Carlsson, F. (2021). The Singleton Fallacy: Why Current Critiques of Language Models Miss the Point. *Frontiers in Artificial Intelligence*, *4*. https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2021.682578

Sarrami-Foroushani, P., Travaglia, J., Debono, D., Clay-Williams, R., & Braithwaite, J. (2015). Scoping Meta-Review: Introducing a New Methodology. *Clinical and Translational Science*, *8*(1), 77–81. https://doi.org/10.1111/cts.12188

Sartori, L., & Theodorou, A. (2022). A sociotechnical perspective for the future of AI: narratives, inequalities, and human control. *Ethics and Information Technology*, *24*(1), 4. https://doi.org/10.1007/s10676-022-09624-3

Satornino, C. B., Du, S., & Grewal, D. (2024). Using artificial intelligence to advance sustainable development in industrial markets: A complex adaptive systems perspective. *Industrial Marketing Management*, *116*, 145–157. https://doi.org/10.1016/j.indmarman.2023.11.011

Savaget, P., Geissdoerfer, M., Kharrazi, A., & Evans, S. (2019). The theoretical foundations of sociotechnical systems change for sustainability: A systematic literature review. *Journal of Cleaner Production*, *206*, 878–892. https://doi.org/10.1016/j.jclepro.2018.09.208

Scharre, P. (2018). *Army of None: Autonomous Weapons and the Future of War*. Norton.

Scott, A. (2021). Difference Between Algorithm and Artificial Intelligence. *Data Science Central*.

Seaver, N. (2019). Knowing Algorithms. In *DigitalSTS*. Princeton university press.

Sen, A. (2004). Capabilities, Lists, and Public Reason: Continuing the Conversation. *Feminist Economics*, *10*(3), 77–80. https://doi.org/10.1080/1354570042000315163

Shaikh, K. (2025, January 29). *Everything you need to know about DeepSeek: The AI disruptor from the east*. Interesting Engineering. https://interestingengineering.com/innovation/everything-about-deepseek-explained

Shaw, I. G. R. (2016). *Predator empire: Drone warfare and full spectrum dominance*. University of Minnesota Press.

Shepherd, L. J. (2021). *"Narrating the Women, Peace and Security Agenda: Logics of Global Governance"* Oxford University Press

Shotter, J. (1993). *Conversational realities: Constructing life through language*. Sage.

Simon, J. (2016). Value-Sensitive Design and Responsible Research and Innovation. In S.-O. Hansson (Ed.), *The Ethics of Technology—Methods and Approaches* (pp. 219–236).

Singler, B. (2019). Existential Hope and Existential Despair in Ai Apocalypticism and Transhumanism. *Zygon®*, *54*(1), 156–176. https://doi.org/10.1111/zygo.12494

Smil, V. (2022). *How the world really works: The science behind how we got here and where we're going*. Viking.

Smith, S. (2005). The contested concept of security. In K. Boothe (Ed.), *Critical Security Studies and World Politics* (pp. 27–62). Lynne Rienner.

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. https://doi.org/10.1108/14779960710846155

Spivak, G. (2015). Can the Subaltern Speak? In *Colonial Discourse and Postcolonial Theory* (pp. 66–111). Routledge.

Squire. (2013). Attuning to mess. In M. B. Salter & E. C. Mutlu (Eds.), *Research Methods in Critical Security Studies* (pp. 37–41). Routledge/Taylor and Francis.

Stirling, A. (2007). "Opening Up" and "Closing Down": Power, Participation, and Pluralism in the Social Appraisal of Technology. *Science, Technology, & Human Values*, *33*(2), 262–294. https://doi.org/10.1177/0162243907311265

Straub, J. (2019). Mutual assured destruction in information, influence and cyber warfare: Comparing, contrasting and combining relevant scenarios. *Technology in Society*, *59*, 101177. https://doi.org/10.1016/j.techsoc.2019.101177

Straube, T. (2019). The black box and its dis/contents: Complications in algorithmic devices research. In M. de Goede, E. Bosma, & P. Pallister-Wilkins (Eds.), *Secrecy and Methods in Security Research* (pp. 175–192). Routledge, Taylor & Francis Group.

Suchman, L. (2020). Algorithmic warfare and the reinvention of accuracy. *Critical Studies on Security*, *8*(2), 175–187. https://doi.org/10.1080/21624887.2020.1760587

Suchman, L. A. (2007). *Human-machine reconfigurations: Plans and situated actions*. Cambridge University Press.

Suleyman, M., & Bhaskar, M. (2023). *The coming wave: AI, power and the twenty-first century's greatest dilemma*. Bodley Head.

Svedin, L. (2011). *Ethics and crisis management*. Information Age Pub.

Taureck, R. (2006). Securitization theory and securitization studies. *Journal of International Relations and Development*, *9*(1), 53–61. https://doi.org/10.1057/palgrave.jird.1800072

Tegmark, Max. (2018). *Life 3.0*. Vintage books.

Theodorou, A., Nieves, J., & Dignum, V. (2022). Good AI for Good: How AI Strategies of the Nordic Countries Address the Sustainable Development Goals. *arXiv Preprint*.

Tobi, H., & Kampen, J. K. (2018). Research design: The methodology for interdisciplinary research framework. *Quality & Quantity*, *52*(3), 1209–1225. https://doi.org/10.1007/s11135-017-0513-8

Transforming Our World : The 2030 Agenda for Sustainable Development, A/RES/70/1 (2015).

Turing, A. (1939). Systems of logic based on ordinals. *Proceedings of the London Mathematical Society*, *2*(45), 161–228.

Turing, A. (1950). Computing Machinery and Intelligence. *Mind*, *59*(236), 433–460.

Turkle, S. (n.d.). "Spinning" Technology. What We Are Not Thinking about When We Are Talking about Computers. In *Technological Visions. The Hopes and Fears that Shape New Technologies* (Vol. 2004). Temple University Press.

UN. (2022). Climate Plans Remain Insufficient: More Ambitious Action Needed Now. *United Nations Climate Change*.

United Nations. (1994). *Human Development Report*. United Nations Development Programme (UNDP).

Vastapuu, L. (2018). Auto-photographing (in)securities Former young female soldiers' post-war struggles in Monrovia 1. In *Visual Security Studies* (pp. 171–188). Routledge.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. ukasz, & Polosukhin, I. (2017). Attention is All you Need. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Verdegem, P. (2024). Dismantling AI capitalism: The commons as an alternative to the power concentration of Big Tech. *AI & SOCIETY*, *39*(2), 727–737. https://doi.org/10.1007/s00146-022-01437-8

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, *11*(1), 233. https://doi.org/10.1038/s41467-019-14108-y

Voorbraak, F. (1996). *Reasoning with uncertainty in AI BT - Reasoning with Uncertainty in Robotics* (L. Dorst, M. van Lambalgen, & F. Voorbraak, Eds.; pp. 52–90). Springer Berlin Heidelberg.

Wajcman, J. (2000). Reflections on Gender and Technology Studies: In What State is the Art? *Social Studies of Science*, *30*(3), 447–464. https://doi.org/10.1177/030631200030003005

Wajcman, J. (2009). Feminist theories of technology. *Cambridge Journal of Economics*, *34*(1), 143–152. https://doi.org/10.1093/cje/ben057

Walia, G. K., Kumar, M., & Gill, S. S. (2024). AI-Empowered Fog/Edge Resource Management for IoT Applications: A Comprehensive Review, Research Challenges, and Future Perspectives. *IEEE Communications Surveys & Tutorials*, *26*(1), 619–669. IEEE Communications Surveys & Tutorials. https://doi.org/10.1109/COMST.2023.3338015

Walker, R. B. J. (1984). *Culture, ideology and world order*. Westview.

Walker, R. B. J. (2016). The Subject of Security. In *Out of Line* (pp. 82–97). Routledge.

Wall, D. S. (2008). Cybercrime and the Culture of Fear: Social science fiction(s) and the production of knowledge about cybercrime. *Information, Communication & Society*, *11*(6), 861–884. https://doi.org/10.1080/13691180802007788

Wallach, W., & Marchant, G. (2019). Toward the Agile and Comprehensive International Governance of AI and Robotics [point of view]. *Proceedings of the IEEE*, *107*(3), 505–508. https://doi.org/10.1109/JPROC.2019.2899422

Watch, H. R. (2020). *Stopping Killer Robots Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control*. Human Rights Watch.

Weber, E. P., & Khademian, A. M. (2010). Wicked problems, knowledge challenges, and collaborative capacity builders in network settings. *IEEE Engineering Management Review*, *38*(3), 57. https://doi.org/10.1109/EMR.2010.5559144

Weber, M. (2011). *Methodology of social sciences*. Transaction Publishers.

Wedeen, L. (2010). Reflections on Ethnographic Work in Political Science. *Annual Review of Political Science*, *13*(Volume 13, 2010), 255–272. https://doi.org/10.1146/annurev.polisci.11.052706.123951

Weitz, N., Carlsen, H., Nilsson, M., & Skånberg, K. (2018). Towards systemic and contextual priority setting for implementing the 2030 Agenda. *Sustainability Science*, *13*(2), 531–548. https://doi.org/10.1007/s11625-017-0470-0

Wibben, A. (2011). *Feminist Security Studies—A narrative approach*. Routledge.

Wibben, A. T. R. (2008). Human Security: Toward an Opening. *Security Dialogue*, *39*(4), 455–462. https://doi.org/10.1177/0967010608094039

Wickson, F., Carew, A. L., & Russell, A. W. (2006). Transdisciplinary research: Characteristics, quandaries and quality. *Futures*, *38*(9), 1046–1059. https://doi.org/10.1016/j.futures.2006.02.011

Wilkinson, C. (2013). Ethnographic methods. In *Critical Approaches to Security*. Routledge.

Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, *109*(1), 121–136.

Winner, L. (2020). *The whale and the reactor: A search for limits in an age of high technology*. University of Chicago Press.

Wooldridge, M. J. (2021). *A brief history of artificial intelligence: What it is, where we are, and where we are going*. Flatiron Books.

Yeh, S. C., Wu, A. W., Yu, H. C., Wu, H. C., Kuo, Y. P., & Chen, P. X. (2021). Public Perception of Artificial Intelligence and Its Connections to the Sustainable Development Goals. *SUSTAINABILITY*, *13*(16). https://doi.org/10.3390/su13169165

Young, M, I. (1990). *Justice and the Politics of Difference*. Princeton university press.

Zanzotto, M. (2019). Viewpoint: Human-in-the-loop artificial intelligence. *Journal of Artificial Intelligence Research*, *64*(1), 243–252. https://doi.org/10.1613/jair.1.11345

Završnik, A. (2020). Criminal justice, artificial intelligence systems, and human rights. *ERA Forum*, *20*(4), 567–583. https://doi.org/10.1007/s12027-020-00602-0

Zehfuss, M. (2018). *War and the Politics of Ethics*. Oxford University Press.

Zehfuss, Maja. (2007). *Wounds of memory: The politics of war in Germany*. Cambridge University Press. http://www.loc.gov/catdir/enhancements/fy0808/2008295173-b.html

Appendix – Interview Guide

A Qualitative Study on Professionals' Imaginaries of AI, Social Impacts, and Security

This study aims to discuss the risks and opportunities associated with AI and its social implications. These aspects have been studied in previous research where security, social sustainability, and AI have been theoretically discussed. Social sustainability is defined as democratic values such as inclusion, justice, and equality. One preliminary finding from that research is that the connection of security, social values, and AI is rarely done in research but is increasingly becoming more important in policy work. Additionally, the study has identified several themes related to the possibilities and potential harms associated with AI development from a social perspective. This talk will be directed from those themes. You have been chosen to participate in this interview because we believe you possess valuable insights and knowledge in this area and can contribute to expanding our understanding of how to approach AI from a social security perspective. Data from this interview will be stored at the Swedish Defence University for ten years after the study is completed, but it will not be used for any purpose other than as specified here. Only the researcher conducting this study will have access to your data, which will not be shared with others. In this study, your name and the organization to which you belong will be anonymized. Participation in this study is voluntary; you can withdraw your consent to participate at any moment. You do not need to explain why you chose not to participate.

Background: Where do you work, and for how long have you held this position? (How do you encounter AI in your profession?)

What is your professional background?

General questions of AI

How would you define AI?

What are your greatest hopes for AI

What are your greatest fears for AI

What do you think of the current debate about AI (ChatGPT)

AI and the good society:

How would you define a good society, and in what way do you believe that AI can fulfill that ideal?

Research has lifted the following positive possibilities:

a). AI can allocate resources in society more fairly while creating new sustainable economic growth methods.

b). Bolster SD and increase the Social Good

c). Strengthen human rights and humanitarian aid

d). Increase security by adding more accurate security measures

Prevent calamities. Find victims of trafficking. However, research also suggests that increased security will enhance primary democratic functions, such as freedom of speech and integrity. How do you perceive this trade-off? Is it worth sacrificing integrity for added security?

e) Speed up technological innovation, leading to prosperity and economic growth that benefits more people, including those in the Global South.

AI and sociotechnical harms:

a). Nudging and Deep fakes can lead to added polarization

"big nudging," citizen scores, and deep fakes. AI has been used in the American election to discredit a competitor. However, there are also positive benefits when nudging can be used to influence people to behave benevolently. How do you view AI and democracy?

b). Trade-off between values of profit and AI for human values.

An obstacle to achieving a "good" society and a human-based AI approach is a capitalist incentive focused on profit and innovation capacity. The risk is that AI development prioritizes profit-making, high-tech solutions when a low-tech alternative may better suit the environment and context.

c). Global South risks lagging behind global digitalization

To develop trustworthy or safe AI, it is vital to have access to high-quality data; without it, AI cannot be effectively trained and used to mitigate the amplification of sociotechnical harms. However, the degree of data availability and accessibility often reflects social, economic, political, and other inequalities. In many development and humanitarian contexts, collecting high-quality data is challenging, which increases the risk that an AI system will produce unfair outcomes.

d). Lack of transparency and traceability is a problem for accountability. (This has been discussed about AI in the security realm and other areas. What happens when AI comes to the wrong conclusion?

e). Built-in biases have consequences for vulnerable groups in society

The technology tends to absorb social biases that work unfavorably for vulnerable groups. (Gehl Sampath, 2021), which can harm the humans affected. This has been especially prominent in the justice system, where an often-cited example is the COMPAS

f) Unpredictable evolution of AI

There is a potential for AI systems to become extremely powerful, generating analytical and predictive insights that increasingly outstrip human capabilities. This means they can be used as replacements for human decision-making, especially when analysis needs to be done rapidly or at a scale.

Concluding Questions:

How do you think ChatGPT and large language models will affect society in the future?

How can we responsibly develop AI? How do we determine which values to program into AI?

Is there any question that you think is important that I have not asked?