TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

# ETHICS AND DEMOCRATIC RESILIENCE IN THE AGE OF PERVASIVE DIGITAL SYSTEMS

## A Rawlsian Approach

Salla Westerstrand

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. E OSA – TOM. 127 | OECONOMICA | TURKU 2025

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

# ETHICS AND DEMOCRATIC RESILIENCE IN THE AGE OF PERVASIVE DIGITAL SYSTEMS

A Rawlsian Approach

Salla Westerstrand

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. E OSA – TOM. 127 | OECONOMICA | TURKU 2025

## University of Turku

Turku School of Economics
Information Systems Science
Doctoral Programme of Turku School of Economics

## Supervised by

University Research Fellow,
Kai K. Kimppa
University of Turku

Senior Researcher, Jani Koskinen
University of Turku

## Reviewed by

Professor Bernd C. Stahl
University of Nottingham, England

Professor Katina Michael
Arizona State University, USA

## Opponent

Professor Katina Michael,
Arizona State University, USA



ISBN 978-952-02-0179-1 (PRINT)
ISBN 978-952-02-0180-7 (PDF)
ISSN 2343-3159 (Print)
ISSN 2343-3167 (Online)
Painosalama, Turku, Finland 2025

*For Rauli.*

UNIVERSITY OF TURKU
Turku School of Economics
Information Systems Science
SALLA WESTERSTRAND: Ethics and Democratic Resilience in the Age of Pervasive Digital Systems: A Rawlsian Approach
Doctoral Dissertation, 228 pp.
Doctoral Programme of Turku School of Economics
May 2025

## ABSTRACT

Pervasive digital systems challenge our understanding of technology on people, societies and the environment. While they come with a promise of making our lives easier, they have raised ethical questions. Pervasive digital systems also influence democratic societies, putting the resilience of liberal democracy to the test. Yet, we lack deep understanding of these impacts, how to study them, and how we – as complex societies – can steer the development of pervasive digital systems towards an ethical direction.

To address this challenge, this dissertation takes a critical stance in order to 1) develop a research approach for studying ethical and societal impacts of pervasive digital systems, 2) shed light on the ethical direction in development and deployment of pervasive digital systems, and 3) increase understanding of how the ethical dimensions influence the fundamentals of democratic societies. The dissertation takes artificial intelligence (AI) as a contemporary example and studies the AI ethics discourse in the context of European AI regulation. It draws from John Rawls's theory of justice as fairness to evaluate the implications of AI from the perspectives of ethics and democratic resilience. It uses critical-political discourse studies as a research approach for empirical analyses.

As a result, this dissertation offers a research approach to study ethical and societal impacts of pervasive digital systems. It introduces a set of ethics guidelines for AI systems, as well as a framework for providers on ethical questions they need to consider beyond EU AI Act compliance. It also proposes a framework for the basic structure of society with which to steer the development of pervasive digital systems towards supporting, rather than eroding, democratic resilience.

This dissertation is thus a significant contribution to both Information Systems (IS) theory and practice. It builds on seminal works in critical IS research and advances the study of sociotechnical ISs, all while offering frameworks for practitioners to take the research into practice. By virtue of grounding the empirical studies on moral and political philosophy, this research strengthens the ethical rigour of IS ethics research and works towards a more holistic understanding of ethics of digital ecosystems.

KEYWORDS: IT Ethics, Artificial Intelligence, AI ethics, John Rawls, justice as fairness, digital democracy

TURUN YLIOPISTO
Turun kauppakorkeakoulu
Tietojärjestelmätiede
SALLA WESTERSTRAND: Ethics and Democratic Resilience in the Age of Pervasive Digital Systems: A Rawlsian Approach
Väitöskirja, 228 s.
Turun kauppakorkeakoulun tohtoriohjelma
Toukokuu 2025

## TIIVISTELMÄ

Läpitunkevat digitaaliset järjestelmät (*pervasive digital systems*) ovat haastaneet yhä useamman miettimään teknologian vaikutuksia ihmisiin, yhteiskuntaan ja ympäristöön. Vaikka ne tuovat mukanaan usein lupauksen helpommasta arjesta ja mukavammasta elämästä, ne nostavat samalla esiin haastavia eettisiä kysymyksiä. Yksilöihin kohdistuvien vaikutusten lisäksi läpitunkevat digitaaliset järjestelmät vaikuttavat yhteiskunnallisiin rakenteisiin ja demokratian kantokykyyn. Emme kuitenkaan täysin ymmärrä näiden tietojärjestelmien vaikutuksia, miten niitä voitaisiin tutkia ja mitata, ja miten vaikutuksiin voitaisiin tosielämässä reagoida.

Tämä tutkimus 1) tuo tutkimusmenetelmällinen lähestymistapa läpitunkevien digitaalisten järjestelmien eettisten ja yhteiskunnallisten vaikutusten tutkimiseen, 2) lisää ymmärrystä läpitunkevien digitaalisten järjestelmien eettisestä suunnasta, sekä 3) lisää tietoa siitä, miten ne vaikuttavat demokraattisten yhteiskuntien rakenteisiin. Väitöskirjassa keskitytään tekoälyjärjestelmiin ja tarkastellaan tekoälyetiikan diskurssia eurooppalaisessa regulaatiossa. Analyysi perustuu John Rawlsin oikeudenmukaisuusteoriaan. Empiiriset osatutkimukset käyttävät Rawlsin teoriaa pohjana, jota vasten tekoälyn voidaan arvioida. Tutkimusmenetelmänä käytetään kriittis-poliittista diskurssintutkimusta.

Tutkimus tuottaa analyyttisen tutkimusmenetelmällisen viitekehyksen, jolla voimme tutkia läpitunkevien digitaalisten järjestelmien eettisiä ja yhteiskunnallisia vaikutuksia. Lisäksi se tuottaa Rawlsin teoriaan perustuvat tekoälyn eettiset periaatteet sekä järjestelmien tarjoajille suunnatun viitekehyksen, joka tuo etiikan osaksi ohjelmistokehitystä. Se tarjoaa myös yhteiskunnan perusrakenteeseen kuuluville instituutioille mallin, jonka avulla teknologiakehitystä voidaan ohjata eettisempään suuntaan tavalla, joka tukee myös demokratian kantokykyä.

Tämä väitöskirja on siten merkittävä kontribuutio tietojärjestelmätieteen teoriaan ja käytäntöön. Se vie eteenpäin merkittäviä kriittisen tietojärjestelmätieteen tutkimuksia ja edistää sosioteknistä tutkimusperinnettä. Samalla se tuo alalle viitekehyksiä, joilla löydöksiä voidaan viedä kohti käytäntöä. Väitöskirjan empiiriset tutkimukset perustuvat moraali- ja yhteiskuntafilosofiaan, minkä ansiosta tutkimus vahvistaa tietojärjestelmätieteen etiikan teoreettista lujuutta ja uskottavuutta sekä vie alaa kohti holistista ymmärrystä digitaalisten ekosysteemien eettisyydestä.

ASIASANAT: IT-etiikka, tekoäly, tekoälyetiikka, John Rawls, oikeudenmukaisuus-teoria, digitaalinen demokratia

# Acknowledgements

This dissertation was written between 2021 and 2025 – years that were marked by a range of disruptions from the Covid-19 pandemic and the wars in Ukraine and Gaza to the introduction of popular generative AI tools that brought profound changes to digital information ecosystems. Many of these events became intertwined, as both the pandemic measures and the rise of automated warfare took advantage of the technological advancements. It was in these circumstances that I conducted this study to deepen our understanding of the impacts of digital developments on people and societies.

In the midst of overlapping political, economic, geopolitical and environmental crises, studying the impacts of AI tools – most of which at the time of writing can still be considered gadgets more than anything else – sometimes felt utterly trivial. Then again, watching how AI systems were deployed in, e.g., battlefields, social security and courts reminded me that we cannot surrender our agency over steering the course of digital development to companies who design algorithms used in making life-changing decisions. As an increasing number of data centres devour natural resources to feed AI products and most individuals become mere drops in the pool of data that some now call the 'new oil', we need to take a step back and observe the impacts of these developments on people's lives across the globe. So here I am, with humble but significant findings that point towards a need for more efforts in ethical digitalisation. Looking back at the journey that led me to this point, I have many people to acknowledge for their support along the way.

The planning of this research started already in 2018, when there was still much less research available about the implications of complex digital technologies such as AI on democratic societies. At the time, I was finalising my master's degree in political science and asked my supervisor's thoughts about doing a PhD on algorithmic decision-making and its impacts on democracy. I got to learn that no political scientists were interested in technology, and no engineers were interested in politics, so it was better to start looking for something else.

I reluctantly took the advice and spent the following years working in roles outside academia. Meanwhile, my curiosity towards the topic continued to grow as I read books, studies and media publications that that made me increasingly puzzled:

why is there such a persistent hype around new technologies while our understanding of the impacts of those technologies on people and societies is still so limited? What makes us think this is the right way to go when there are indications that the overall impact on people's lives could be negative, if not disastrous? As 2021 approached, encouraged by several inspiring discussions with friends and family, I made up my mind: it was my time to dig into the ethical and societal impacts of these systems that had become ever more pervasive in our lives.

I knew of very few research groups with a focus on such issues, so I consider myself lucky to have found my way to Turku School of Economics and the Future Ethics research group, where I was warmly welcomed from the day one to contribute to the shared hope for more ethical information systems development. Coming to the field of information systems with a background in political science and French language meant that I was faced with a challenge of adopting a new field of study while bringing in new perspectives from my *alma mater*. Although I was already well versed in multidisciplinary studies, I was suddenly in a field about which I thought I knew practically nothing – a prospect that in the beginning felt intimidating.

My worries evaporated quickly as I was fully supported by my research group in these efforts. I am especially thankful for my supervisors: Kai Kimppa, who saw the value in my originally vague research proposal, provided me with a well-measured challenge and supported my journey with thorough comments and advice that contributed to the development of my thinking and argumentation; and Jani Koskinen, whose unconditional, persistent encouragement pushed me to keep the bar high and made me feel like a fully established member of the academic community since the very beginning. You kept me involved in research even in times when the circumstances forced me to accept a full-time job outside academia. I am also thankful to all the other Future Ethics research group members with whom I got to work over the years. I cherish the memory of my first academic conference that took place only two months after I started my PhD – you happily put me to chair one session with a 2-day notice. Even though I had no idea what a session chair does, you all were somehow confident that I would do fine. This kind of support ensured that I never felt like 'just a PhD student' whose contribution is less valuable than that of an experienced researcher. I consider that to be one of the key factors that made my PhD journey so fulfilling and built my confidence as a researcher up to this day. I want to particularly thank Minna M. Rantanen, whose guidance and perspectives were invaluable in my path towards finalising this dissertation, and Antti Tuomisto, whose research projects ensured that I had a solid start for my research career and also my PhD – TTDLoikka taught me a lot but most importantly made it possible for me to jump full-time into research.

I also want to thank professors Katina Michael and Bernd C. Stahl for your pre-examination and valuable comments to this dissertation, which prepared me for the defence and yielded many ideas for further research. Special thanks also to Theodore Lechterman for bringing me to the ECPR General Conference in Dublin in August 2024, which ended up marking a pivotal point in how my dissertation eventually came together. Thank you, Ted, also for your comments to the last one of the articles included in this dissertation – I keep returning to those often, wondering about ways to keep pushing my understanding around the relationship between technology, democracy and justice. That work certainly continues in the future.

I would also like to thank my colleagues at Solita, who, during the final year of conducting research for this dissertation, cultivated my thinking and reflection of the practical implications of my research. Working hands-on in data and AI teams added to the meaningfulness of this research and my hope that the work will have an impact in the industry. Special thanks to those of you who understood the value of academic work and the need for sometimes fuzzy, complex and unfinished ethical deliberation as a driver for change – Antti Rannisto and Satu Korhonen in the forefront. You were always there to challenge existing ideas, offer fresh perspectives and bring up new theories I had not thought of before. You are truly amazing colleagues.

I also want to thank Karolina Drobotowicz and Ana Paula Gonzalez Torres for insightful discussions around ethical and societal implications of AI. Visiting you at Aalto was always a pleasure. You contributed to my thinking and challenged my assumptions in a way that allowed me to grow as a researcher. I hope we will have many more discussions and opportunities to collaborate in the years to come.

I want to thank Turku Foundation of Business Education, Matti Koivuranta foundation and the Information Systems department of Turku School of Economics for conference travel funding.

I also want to thank my family – my mother, father and all three siblings – for your support. I admire your patience in listening to the hurdles of working in academia, and your unconditional joy and support over successes, and empathy over challenges. Sharing this experience with you made it ever more meaningful. Having a strong support network was invaluable in times of uncertainty with research funding, and having an outsider perspective to the whole PhD process was often sobering and comforting. I am lucky to have you all in my life.

Undoubtedly the biggest gratitude I feel for my beloved partner-in-life, Rauli Westerstrand, who took me into his universe, fully believed in me, and inspired me to act upon the persisting dream of becoming a researcher. This dissertation would have never been started nor finished without you. Your infinite support and rogue but so very acute observations about the world have elevated me and my research since day one. Together we have fought through hardships, uncertainty, celebrations of accepted papers and disappointments of rejections. I cherish the countless long

nights, intense philosophical discussions (that I am tempted to call debates due to their intensity but cannot really as we mainly tended to agree), fuelled by your excellently refined taste in affordable red wine. You made sure my doubts about my skills never grew disproportionate – you kept us going even when I felt like giving up. With you, I have climbed higher and dived deeper than I ever thought possible – both figuratively and literally. I can see us still talking about the complexities of liberal capitalism, technology, ethics and democracy in our rocking chairs decades from now, as that is who we are, that is our process. To you I dedicate the findings of this work and all deliberations to come.

Needless to say, I am extremely privileged to have had all this support during my PhD journey. So privileged indeed that there would have been many more to recognise for their contribution to my thinking. Each encounter in conferences, seminars, social gatherings and community events left a mark that made me think about the impacts of digitalisation from a fresh perspective. It leaves me quite speechless. This feeling of gratitude grows ever greater as I steer my attention towards the future and imagine the opportunities ahead. For now, I hope this dissertation brings you, the reader, as many inspiring moments as it did to me.

26 April 2025
*Salla Westestrand*

## SALLA WESTERSTRAND

Westerstrand (MSSc., MA) is an information systems researcher with background in political science and language studies. She focuses on ethical and societal implications of information systems, embracing multidisciplinarity and academia-industry collaboration. At the time of writing, she works as AI designer, researcher, board member, consultant and advisor – bringing ethics into practice to encourage ethical and societally sustainable digitalisation.

# Table of Contents

## Tables



## Figures

# List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Westerstrand, S. Ethics in the intersection of AI and democracy: The AIDEM Framework. 2023; *ECIS Research Papers*, https://aisel.aisnet.org/ecis2023_rp/321/.

II Westerstrand, S. Reconstructing AI Ethics Principles: Rawlsian Ethics of Artificial Intelligence. 2024; *Science and Engineering Ethics.* 30(46). https://doi.org/10.1007/s11948-024-00507-y.

III Westerstrand, S. When Information Systems Go Political: A Research Approach for Political IS Discourse; *Unpublished manuscript, under review in European Journal of Information Systems.*

IV Westerstrand, S. Fairness in AI Systems Development: Beyond EU AI Act Compliance. 2025; IN Papatheocharous, E., Farshidi, S., Jansen, S., Hyrynsalmi, S. (eds)*, Software Business. ICSOB 2024. Lecture Notes in Business Information Processing,* vol 539. Springer, Cham. https://doi.org/10.1007/978-3-031-85849-9_9.

V Westerstrand, S. Towards Just Democracies in the Age of Pervasive Digital Systems – A Rawlsian Approach. *Unpublished manuscript, under review in AI & Society.*

# Key Concepts and Definitions

| | |
|---|---|
| Applied ethics | Application of ethics theories to real-life contexts to solve moral dilemmas, and/or to increase understanding of the moral nature of the phenomenon under scrutiny. |
| Artificial Intelligence | A discursive phenomenon revolving mainly around algorithmic systems that function with a certain level of autonomy, i.e., without a technical need for constant human involvement, developed using techniques such as machine learning that allow the system to evolve during its lifecycle based on external input (e.g., computer vision or natural language) or training data. |
| Democratic resilience | "The ability of a democratic system, its institutions, political actors, and citizens to prevent or react to external and internal challenges, stresses, and assaults through one or more of the three potential reactions: to withstand without changes, to adapt through internal changes, and to recover without losing the democratic character of its regime and its constitutive core institutions, organizations, and processes" (Merkel & Lührmann, 2021, p. 874). |
| Discourse | A linguistic entity constructed in speech acts that are shaped by the context. |
| Ethics | The branch of philosophy that studies what is morally good and bad, and right and wrong. Also known as moral philosophy. |
| Pervasive digital systems | Systems that are general-purpose, scalable, unpredictable and complex, and come with low transaction costs, and thus "penetrate human life, experience, products, business processes, and civic society" (Grover & Lyytinen, 2023, p. 47). |

# 1 Introduction

New generations of digital technologies have challenged our understanding of the role of technology in societal change. Digital platforms seem to profoundly influence our everyday decisions and societal structures (Susskind, 2022; Muldoon & Raekstad, 2023; Coeckelbergh, 2024c). We hear promises that Information Systems (ISs) utilising artificial intelligence (AI) techniques make our lives easier by facilitating mundane tasks, providing better health (e.g., Bates et al., 2021; Javaid et al., 2023) and more efficient transport (Du et al., 2023). Meanwhile, some characteristics of AI systems have been shown to tamper with freedom and human autonomy (e.g., Formosa, 2021; Prunkl, 2024), justice and fairness (Douglas, 2015; Franke, 2021; Gabriel, 2022; P. Hacker, 2018; Heidari et al., 2019; Mehrabi et al., 2022; Mitchell et al., 2021) and trust towards democratic institutions (Chesney & Citron, 2019; Manheim & Kaplan, 2019; Paterson & Hanley, 2020). Generative AI systems are being used to produce convincingly real-looking but fake content online that has been shown to affect our political opinion-formation and freedom of elections (Coeckelbergh, 2024c; Jungherr & Schroeder, 2023; Łabuz & Nehring, 2024; Mainz et al., 2024).

Still, the hype around generative AI applications is stronger than ever with new models and APIs rushing into markets, bringing forth concerns around the ethics of AI hype (Westerstrand et al., 2024). Simultaneously, the wilderness of the metaverse has raised questions of monetary, social, and political motives of actors that aim to define it (Dolata & Schwabe, 2023) and provoked discussion about the ontological assumptions about those virtual realities and their ethical and political implications (Coeckelbergh, 2024a).

Whereas the introduction of new technologies that shape our societies is not a new phenomenon, the increasing complexity (Benbya et al., 2020) and characteristics such as general-purpose use, scalability, unpredictability and low transaction costs make digital systems ever more pervasive in our everyday life (Grover & Lyytinen, 2023). We are faced by the *pervasive digital*, as Grover and Lyytinen (2023) have dubbed it, which provokes questions about who creates change and decides its direction, as their uses “penetrate human life, experience, products, business processes, and civic society” (p. 47).

Among pervasive digital systems, a set of technologies addressed by the name AI has received exceptional attention, bringing forth both opportunities and challenges. As a response, we are witnessing an emergence of new research areas – if not disciplines – such as AI ethics (see, e.g., Stahl, 2022) and AI governance (e.g., Birkstedt et al., 2023; Ibáñez & Olmeda, 2022; Mäntymäki, Minkkinen, Birkstedt, et al., 2023; Mäntymäki, Minkkinen, Zimmer, et al., 2023; Morley et al., 2021, 2023), which aim to bring clarity into the implications of AI systems and how to govern them. A plethora of principles and guidelines have been proposed by governments, industry organisations and NGOs to guide the way we develop and deploy AI systems (Franzke, 2022; Hagendorff, 2020; Jobin et al., 2019), and regulatory initiatives such as the European AI regulation, (EU AI Act) have added to the pressure of bringing principles into measurable actions.

Meanwhile, it seems that we are still perplexed by the impacts – both positive and negative – of pervasive digital systems on people, societies and the environment. What kind of ethical implications could these new systems have? Do they shape our societal structures in ways we do not fully understand? How do ISs shape the power structures of our societies? What does that mean for human rights, such as freedom and autonomy? Whereas the initial attempts have already pointed out some implications and promising approaches for governance (for a literature review, see Birkstedt et al., 2023), the rapid advances in AI have pushed us towards perhaps premature initiatives to operationalise ethics: there is a persisting lack of ethical justifications in AI ethics principles and their operationalisation (Bleher & Braun, 2023; Franzke, 2022), which risks resulting in guardrails and tools that are "either inappropriate, meaningless, or merely an end in themselves" (Bleher & Braun, 2023, p. 10). It thus seems that for the IS research community, the ongoing developments call for new, innovative, and holistic theorising and knowledge creation amongst academics (Grover & Lyytinen, 2023). For practitioners and academics to design and deploy ethical ISs and governance models that support the resilience of democratic societies, we still need to build the foundations that ensure these guardrails are taking us to a desired direction. Furthermore, to define that direction, public deliberation needs to be fostered to evaluate what the common good looks like in the age of pervasive digital systems (Coeckelbergh, 2024b).

This dissertation contributes to the need for deeper understanding of the ethical and societal implications of pervasive digital systems and how to study them. Taking a multidisciplinary approach that combines IS research, ethics and political theory, this study sheds light on ethical dimensions of the development of pervasive digital systems and the implications of those dimensions on democratic societies. I adopt a critical stance and use AI as an example of a pervasive digital system that comes with societal and ethical impacts. I draw from John Rawls's theory of justice as fairness to study the ongoing ethics discourse around AI and how it impacts Rawlsian

ideal of just democracy, as well as its resilience against forces that erode its democratic foundations. I also offer a research perspective that allows us to keep studying IS phenomena that are increasingly pervasive. Therefore, I seek response to the following research questions:

- RQ 1: Methodologically, how could we better increase the understanding of sociotechnical dimensions of pervasive digital systems?
- RQ 2: From the perspective of John Rawls's theory of justice, what kinds of ethical dimensions can we distinguish in the ongoing AI ethics discourse?
- RQ 3: In the light of the response to RQ2, what kind of implications do pervasive digital systems have on democratic resilience?

These questions are addressed in articles as indicated in Table 1.

**Table 1.** Articles belonging to this dissertation, the corresponding research questions, data and methodologies used.

| Article n° | Title | Research question | Data | Methodology |
|---|---|---|---|---|
| I | Ethics in the intersection of AI and democracy: the AIDEM framework. | **RQ 1** RQ2 | Spanish AI strategy | Conceptual + Critical Discourse Analysis |
| II | Reconstructing AI Ethics Principles: A Rawlsian Approach | RQ 2 | - | Conceptual |
| III | When Information Systems Go Political: A Research Approach for Political IS Discourse | **RQ1** RQ 2 | EU AI Act | Critical-Political Discourse Studies |
| IV | Fairness in AI Systems Development: Beyond EU AI Act Compliance | RQ 2 | EU AI Act | Critical-Political Discourse Studies |
| V | Towards Just Democracies in the Age of Pervasive Digital Systems – A Rawlsian Approach | RQ 3 | - | Conceptual |

The dissertation thus starts with drafting an initial analytical framework that conceptualises the impacts of pervasive digital systems – AI as an example – on democratic societies through an ethical lens (I). It then explores the moral philosophical foundations of AI ethics principles that guide the efforts in AI development, governance and policy, proposing a set of principles grounded in Rawls's theory of justice as fairness (II). It then focuses on the development of the

research approach to ensure that we have tools for rigorous analysis of politically loaded IS discourses (III). Using the framework and the methodology developed in articles I and III, it then provides an analysis of the ongoing AI discourse by focusing on the European AI regulation – the EU AI Act – and what kind of premise it provides for ethical AI development (IV). The final article (V) builds on knowledge in the previous articles and AI ethics literature, and presents an analysis of what kinds of implications the current AI ethics discourse has for the fundamentals of democracy through the lens of Rawls's basic structure of society. From this background, in this dissertation, I propose critique towards the status quo and offer suggestions for steering the development and deployment of ISs to a more ethically and societally sustainable direction.

Before embarking on a more detailed description of the background and the theoretical foundations of this research, some essential definitions and clarifications are needed. Using AI as an example of pervasive digital in several articles requires a definition of what is included in the scope of AI systems. I do not intend to contribute to the ongoing debate around the best possible definition of the term artificial intelligence as a specific technology. Delimiting the scope only to very specific technical solutions risks losing a crucial aspect that contributes to the ethical and societal implications of these technologies: regardless of the technical details, the ways in which we talk about AI, the narratives and imaginaries we create and the language we use to describe these systems and their capacities, limits and characteristics can be seen as constructive of the ISs themselves (see, e.g., Deetz, 1996; Orlikowski, 1992; Swanson & Ramiller, 1997).

Therefore, this dissertation looks at AI as a discursive phenomenon revolving mainly around algorithmic systems that function with a certain level of autonomy, i.e., without a technical need for constant human involvement, developed using techniques such as machine learning that allow the system to evolve during its lifecycle based on external input (e.g., computer vision or natural language) or training data. AI systems are here used as an example of pervasive digital systems in the meaning described by Grover and Lyytinen (2023) discussed above. Following the perspective adopted also in several European regulatory initiatives, the definition thus emphasises the autonomy and adaptability of the system and the way in which it has been used. This definition is thus broad enough to include rather simple recommendation algorithms and systems with mere assistive function in decision-making, as they can be seen to play an essential role in the discourse that further shapes the course of AI development and deployment. Hence, despite the obvious discomforts that come with inclusive and vague definitions, in the context of this research, this has been a deliberate choice that serves the aims of the research in seeking answers to the research questions.

This dissertation focuses on applications of artificial narrow intelligence (ANI) and does not engage in the discussion around the alleged possibility of emergence of artificial general intelligence (AGI) or superintelligence (Bostrom, 2017). This choice is informed by the fact that at the time of writing, such systems do not exist. There is also a vast body of research in fields such as philosophy of mind that questions the possibility of such systems ever being brought to life (for an extensive analysis, see Bennett & Hacker, 2021). As meaningfully engaging in this discussion would merit its own dissertation, it suffices here to state this limitation. Yet, it is essential to recognise that this presupposition means this dissertation focuses on systems based on currently available AI-based ISs that already influence people's lives and societal structures. Without neglecting the importance of exploring the technological limits of creating artificially "intelligent" machines, considering the hypothetical if not speculative nature of the discourse surrounding such systems, I argue that the present approach responds better to the needs of today's societies and those of the foreseeable future. This offers a more rigorous basis for strengthening our understanding of the implications of pervasive digital systems on people and societies – including technologies we are yet to see.

When it comes to evaluating the impacts of AI on democracy, I focus on democratic *resilience*. As Merkel and Lührmann (2021) discuss, there is an abundant literature studying different forms of democratic decline, while fewer contributions are made towards studying the success factors that prevent democracies from eroding, backsliding or declining. This tendency seems to have taken a turn, however, as literature reviews indicate a growing scholarly interest towards a study of democratic resilience (Holloway & Manwaring, 2023). I rely here on Merkel and Lührmann's definition, according to which democratic resilience is "the ability of a political regime to prevent or react to challenges without losing its democratic character" (Merkel and Lührmann, 2021, p. 872). This definition avoids many pitfalls found by Holloway and Manwaring (2023) in earlier attempts to conceptualise democratic resilience, such as unambiguity in definitions and repackaging existing concepts under the name of resilience. In this dissertation, looking at the different levels of a political system – political community, institutions, actors, citizens (Merkel & Lührmann, 2021) – I focus on institutions that Rawls conceptualises as the *basic structure of society* to evaluate whether the current AI ethics discourse has an impact on its resilience in securing just background conditions for democratic societies.

Lastly, some might question the choice of drawing from both ethics and democratic theory in a dissertation situated in the field of IS research. Considering the breadth of each field of study alone, it could have been justified to focus only on one at a time. However, since the early stages of this research, it became apparent that the questions around impacts of pervasive digital systems discussed by, e.g.,

information systems scientists, ethicists and political scientists, were intertwined. Consequently, I argue that a thorough understanding thereof benefits from a holistic, multidisciplinary perspective. It needs to be recognised that this choice means a limited range of ethics perspectives covered in the empirical analyses of ethical implications, as well as a limited set of perspectives to democratic theory – i.e., the focus on Rawls's theory of justice alone. Whereas this can be seen as perhaps the biggest limitation of this dissertation, it also enables its biggest contribution to the field. This choice has allowed me to reveal interdependencies and connectedness that would have otherwise been underdiscussed. In an environment where academic competition drives us towards ever narrower perspectives to complex, convoluted topics, I argue that it is essential to create links between phenomena and fields of research, even if it means that there remain gaps for further research. Hence, I hope this dissertation serves as an inspiration to others, too, to start filling in the gaps – one perspective at a time – to keep completing the picture the contours of which I draw in this dissertation.

In what follows, Chapter 2 discusses the background and literature around ethical and societal implications of ISs and their governance that motivate this research, focusing on AI systems. Chapter 3 unpacks the ontological, epistemological and methodological choices of the dissertation. Chapter 4 discusses the theoretical background in John Rawls's theory of justice and justifies the choice of taking it as a starting point for this dissertation. Chapter 5 offers a synopsis of the articles included in this study and a synthesis of their findings, and Chapter 6 discusses the implications of the results from the perspectives of ethics discourse and democratic resilience. Finally, Chapter 7 ends with conclusions.

# 2 Background: Ethical Information Systems and Democratic Societies

## 2.1 IT ethics, AI ethics, and beyond

Ethical implications of ISs have been a recurring topic in various fields of research, ranging from Information Systems, Science and Technology Studies, and Computer Science to Sociology and Communications Science (Kazim & Koshiyama, 2021; Stahl, 2022). In IS research, several academics have contributed with their seminal works to the development of IT ethics (e.g., Chiasson et al., 2018; Hirschheim et al., 1995; Mingers & Walsham, 2010; Mumford, 1983, 1998, 2003; Porra & Hirschheim, 2007). This dissertation builds on this body of knowledge and shifts the focus towards ethical questions related to pervasive digital systems. At the time of writing, much of the academic interest towards ethical implications of IT revolves around artificial intelligence (AI), which has led to an emergence a field of study titled AI ethics (see, e.g., Stahl, 2022). Due to this heavy emphasis on AI systems in research efforts, the developments in AI ethics have been used as a starting point to understand how the impacts of pervasive digital systems are being studied and where the discussion might be heading next.

In the realm of moral philosophy, AI ethics is situated under *applied ethics*, which draws from normative theories and applies their justifications for what is morally good and bad, or right and wrong, in a particular situation involving AI systems. Whereas the metaethical positions of these perspectives vary (see, e.g., Miller, 2003), most theories assume that we can define key elements for morally righteous action, which can guide us in developing and deploying AI systems that are in line with those moral philosophical assumptions. Whereas metaethics typically feeds normative ethics, which again feeds applied ethics, observations from the field sometimes also shape normative theories and thus also give feedback to metaethical considerations, making the relationship between different layers interactive (see Figure 1).

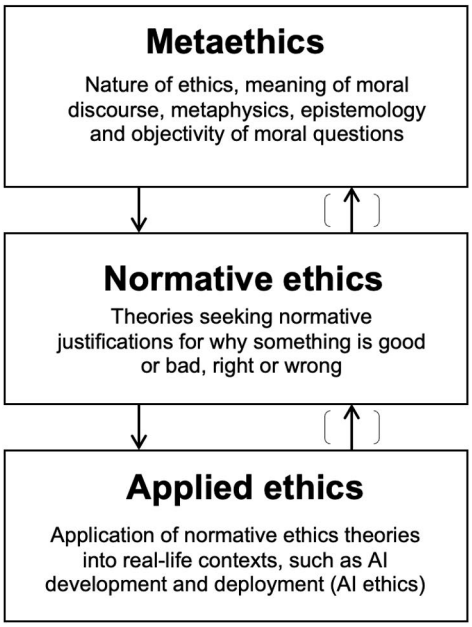


**Figure 1.** Relationship between different ethics perspectives.

From a temporal perspective, recent developments that build on the aforementioned research traditions in IT ethics can be characterised in three waves: 1) principle-based AI ethics, 2) operationalisation of AI ethics principles, and 3) ethics of digital ecosystems. These partially overlapping waves of research are illustrated in Figure 2.

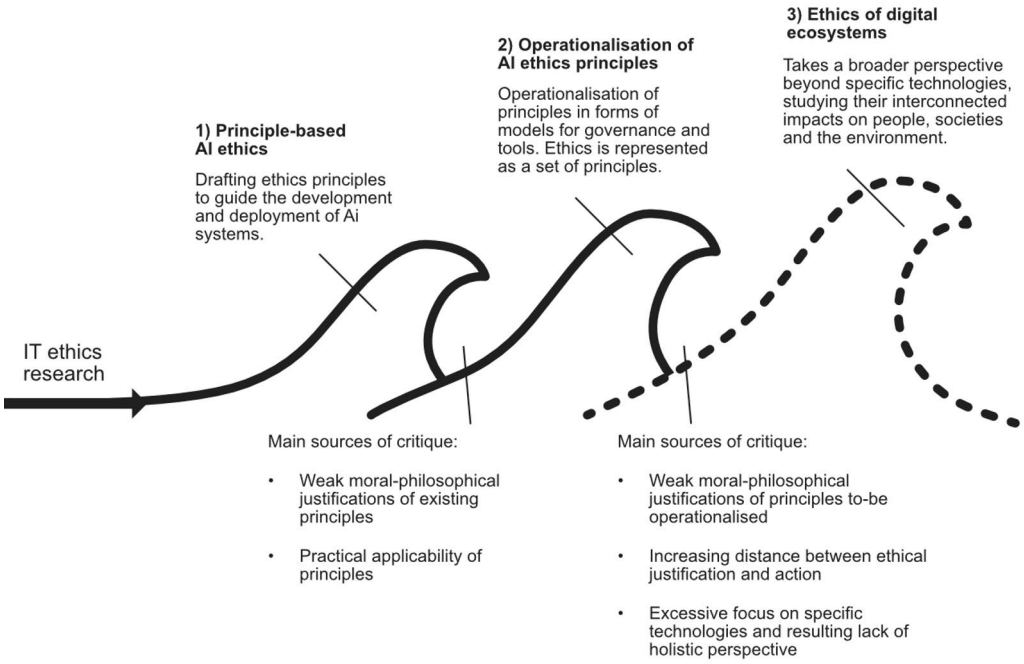


**Figure 2.** Three waves in research of ethics of pervasive digital systems.

**The first wave** has mainly concentrated on drafting ethics principles to guide the development and deployment of AI systems (for reviews, see, e.g., Franzke, 2022; Hagendorff, 2020; Jobin et al., 2019). This perspective – also known as the *principle-based approach* – has been criticised by scholars for its tendency to rely too heavily on pre-defined principles to the detriment of active ethical reflection (e.g., Bleher & Braun, 2023; Mittelstadt, 2019; Rességuier & Rodrigues, 2020).

Critique has been directed towards applying bioethics principles to AI without necessarily justifying why they would provide the most appropriate ethical framework also for studying ISs that are used in varying contexts that all can have their own professional codes and practices (Mittelstadt, 2019). For example, Floridi et al. (2018) introduced a framework titled *AI4People* that is rooted in the four principles of bioethics – autonomy, justice, beneficence and maleficence – and an added principle of explicability. Whereas using these principles can be justified when AI is used in medical contexts (Beil et al., 2019; Grote, 2022), applying them to AI in general with little moral-philosophical justification (Hermann, 2022; Schiaffonati, 2022) is questionable. Therefore, although principles in general play a role in ethical IS development (Seger, 2022) and form an integral part of traditions in classical ethics as well as applied ethics (Miller, 2003), as argued by, e.g., Bleher

& Braun (2023) and Rességuier & Rodrigues (2020), reducing ethics into *mere* principles instead of presenting it as an active process for proactive ethical decision-making risks diluting ethics into checklists that give no guarantee of encouraging ethical action.

As a response to the critique received by the principle-based approach, we can distinguish **a second wave** of AI ethics revolving around operationalisation of principles – the one we are currently witnessing. Several studies have proposed frameworks that offer ways to operationalise the principles in IS practice (Ibáñez & Olmeda, 2022; Krijger, 2022; Morley et al., 2021, 2023), and proposed models for governance that recognise ethics principles as one element steering the development towards allegedly ethical direction (Mäntymäki et al., 2022; Mäntymäki, et al., 2023; Mäntymäki et al., 2023). However, when focusing only on operationalising pre-defined principles that often lack ethical justifications (Franzke, 2022; Koniakou, 2023), we quickly find ourselves struggling with the initial problem arising from the lack of ethics in AI ethics principles (Bleher & Braun, 2023). Moreover, when ethics is seen as a separate, top-down block in AI governance that needs to be operationalised in a form of processes and tasks (see, e.g., Mäntymäki et al., 2023), the distance between the moral philosophical justifications and the actual actions grow to an extent that risks disconnecting the IS practice from its ethical justifications. It thus seems that we are currently in a state where the role of ethics in AI ethics leaves room for improvement.

Parallel to the development of principles and models for operationalisation, research studying the ethical implications of ISs has been increasing in volume. It has revealed implications to, e.g., freedom and human autonomy (e.g., Formosa, 2021; Prunkl, 2024), justice and fairness (Douglas, 2015; Franke, 2021; Gabriel, 2022; P. Hacker, 2018; Heidari et al., 2019; Mehrabi et al., 2022; Mitchell et al., 2021), and human flourishing (Stahl, 2021b; Stahl et al., 2021). An increasing number of pieces of news and research is questioning the environmental sustainability of complex algorithmic systems that seem to require an overwhelming amount of resources to be trained and maintained (Bender et al., 2021; Brevini, 2022; Dhar, 2020; Markelius et al., 2024; Ren & Wierman, 2024; Rillig et al., 2023). Many studies have been conducted on topics and fields closely related to ethics, such as approaches concentrating on human rights (Hunkenschroer & Kriebitz, 2023; Livingston & Risse, 2019; Pizzi et al., 2020; Risse, 2019; Romero Moreno, 2024) and value-sensitive design (Sadek et al., 2023; Umbrello et al., 2021; Umbrello & van de Poel, 2021). Characteristic to an emerging field, AI ethics is still finding its contours, which is why the definition of AI ethics is still fuzzy (Raab, 2020). We are also still in the process of understanding who is an AI ethicist (Cocchiaro et al., 2024) and what does it mean to apply moral philosophy rigorously in academic research and practice. It is thus clear that when talking about AI ethics, we are not talking

about a unified field of study that shares common roots and theory basis. In fact, we are not necessarily always talking about ethics at all, which brings an added layer of complexity into building a knowledge basis for understanding the ethical implications of pervasive digital systems on people, societies and the environment.

The importance of drawing from ethics as moral philosophy – both in the context of AI but also in IS research in general – has been demonstrated by, e.g., Stahl and Eke (2024), who used applied ethics methods to better understand the implications ChatGPT on people and societies, showing the value of rigour and ethics-based methods in adding systematicity and depth to their analysis. Sarker et al. (2019) call for ethical reflection in IS research to ensure the longevity of the field and preservation of cohesion, rooted into sociotechnical perspectives. Ethics has been called for in the above-mentioned neighbouring fields of value-sensitive design (Jacobs & Huldtgren, 2021) and human rights (Stahl et al., 2022), which highlights the potential of ethics to add to the rigour of study of impacts of pervasive digital systems such as AI.

There is thus a need for deeper understanding of the *ethics* of AI – the moral philosophical justifications that allow us to understand why some use cases are problematic in the first place, and why following certain set of principles would be morally justified. Relying on moral philosophy can also help us understand how to navigate the inevitable value conflicts that arise when applying principles in IS practice, which helps us to make ethically informed and potentially more transparent trade-offs (for further discussion on ethical AI principles, see Article II).

Looking at the programmes of IT ethics conferences such as ETHICOMP and Tethics, as well as some recent Special Issue themes in the Senior Scholars' list of Premier Journals, a.k.a., the Basket of Eleven IS journals[1], topics around AI systems and their impacts on people and societies have gained significant interest in research communities. There is, however, also a recognised hype around AI systems (which has yielded its own special issue in Springer's AI and Ethics journal[2]), which is one likely reason for why much of the scarce research resources dedicated to the study of the impacts of technology have now been directed to AI-focused research.

[1] For example, topics of Generative AI and DSR particularly discussing AI in Decision Support systems (https://www.sciencedirect.com/journal/decision-support-systems/about/call-for-papers); The Dark Side of Analytics and Artificial Intelligence in European Journal of Information Systems (https://www.tandfonline.com/toc/tjis20/31/6); Humans, Algorithms, and Augmented Intelligence in Information Systems Research (https://pubsonline.informs.org/toc/isre/32/3); several special issues on Generative AI and beyond in Journal of the AIS (https://aisel.aisnet.org/jais/specialissues.html); Generative AI in Journal of MIS (https://www.jmis-web.org/cfps/JMIS_SI_CfP_Generative_AI.pdf).

[2] Special issue titled "The Ethical Implications of AI Hype: Examining the overinflation and misrepresentation of AI capabilities and performance" in AI and Ethics: https://link.springer.com/collections/hcehbeagfd.

However, no hype lasts forever, which is why we seem to be entering what I call **the third wave** in the ethics of pervasive digital systems. Compared to previous waves, the third wave takes a broader perspective to the range of technologies and their interconnected impacts on people, societies and the environment. Although gaining deeper understanding of the impacts of AI is vital in times when organisations are quick to implement new systems without necessarily considering the whole range of their impacts, it needs to be recognised that we need to look beyond AI and discern patterns that make these systems so impactful.

The third wave has its roots in research conducted parallel to the earlier waves and thus is not something entirely new. Rather, it is a direction that has been in the making since the focus on specific technologies prior to the popularisation of AI took place. Stahl (2021a, 2022) suggests moving beyond technology-specific approaches such as "computer" ethics or ethics of "AI" towards an ecosystem approach to look at ethical and normative dimensions of digital ecosystems. Stahl describes the advantages of such approach as follows:

> "The benefit of using such a digital ecosystems-based approach is that it moves away from a particular technology and opens the view to the way in which technical developments interact with social developments, which broadens the view to encompass application areas, social structures societal environments as well as technical affordances. Actual ethical concerns are affected by all of these different factors and the dynamics of their relationships." (Stahl, 2022, p. 72.)

Therefore, rather than always looking at the impacts of one specific technology in isolation, we would benefit from directing more focus towards understanding how ethical issues arise in digital ecosystems and how they and their root causes are interconnected. In this perspective, moving from specific technologies such as AI systems towards a concept of pervasive digital systems by Grover and Lyytinen (2023) seems particularly fruitful. It recognises the specificities of pervasive digital systems – their complexity, ubiquity, scalability and sociotechnical and general-purpose nature – but also their relationship with the surrounding context. It facilitates the study of systems that consist of many different technologies that play into those characteristics. For instance, social media platforms are arguably pervasive digital systems that can be studied from many perspectives: the AI and algorithmic components that recommend contents, the impacts of systems used to produce (fake) contents on those platforms, as well as infrastructural choices for interaction and communication, moderation, censure, and more. All these features could be studied separately but only looking at their interconnected impacts we start to see the holistic impacts they play in society and people's lives (for more discussion on the complexities of digital platforms, see Muldoon, 2022).

One perhaps obvious barrier to this approach arises from the nature of pervasive digital systems themselves: they are complex, general-purpose, ubiquitous, scalable and sociotechnical (Grover & Lyytinen, 2023). All these dimensions challenge researchers when defining the scope of research and the limits of a digital ecosystem. In the age of quantified academics that highlights speed and quantity of publishing research (Koskinen et al., 2024), grasping such entities can seem frustrating. In this dissertation, my aim has been to take on the challenge and broaden the perspective by using the discourse around AI systems as an example of how we can analyse ethical and societal dimensions of pervasive digital systems when they span across geographic and sectorial borders. My work is thus situated in between the waves 2 and 3 illustrated in Figure 2: much of the research included in this dissertation can be seen to fall into the realm of AI ethics, yet the conclusions they form look forwards and seek for patterns beyond AI systems and focus on the impacts of pervasive digital systems. I argue that this approach allows us to root the observations in contemporary phenomena revolving around AI systems and thus validate them rather than rely on mere speculation of potential future technologies, all while highlighting impacts that are not dependent on the technical composition of AI systems but rather their role in people's live and societal change. This also opens the pathway towards informed reasoning of the impacts of technologies we are yet to witness.

The ecosystem perspective is also visible in this dissertation through the emphasis on societal impacts of AI systems as a companion to ethical analyses. As I argue in Article I, ethics can help us understand the fundamental mechanisms of AI's impacts on democracy, which is discussed in more depth below.

## 2.2 Towards ethical digital democracies?

Parallel to the literature looking at ethical implications of ISs, scholars from fields of political science and IS research alike have begun to pay attention to societal implications of digital systems. Technologies such as the internet (Farrell, 2012), social media (Loader & Mercea, 2012; Persily et al., 2020; Price, 2013), electronic voting (Heimo et al., 2010) and election aid applications have been extensively studied over the years, all of which are generally grouped under the umbrella of *digital democracy* (Congge et al., 2023; K. L. Hacker & Dijk, 2000; Helbing, 2021; Weinhardt et al., 2024).

Meanwhile, despite the recognised importance of the role of algorithmic decision-making in democratic contexts (e.g., Nemitz, 2018), AI has received much less interest until the release of popular generative AI platforms such as ChatGPT and Midjourney. Some attempts have been made to unpack the relationship between the ongoing AI development and democratic forms of governance. For example,

Mark Coeckelbergh (2024c, 2024b) discusses negative implications of AI on democratic principles – especially knowledge and trust – and suggests as a response stronger public deliberation and renewal of both technology and our political systems aiming towards common good. He argues that the current system leads to producing a "small technocratic elite that rules a mass of angry citizens who rightly complain that they are not heard but fail to see that" (Coeckelbergh, 2024b). Doing so, Coeckelbergh reveals several aspects of democratic change where technology plays a role. Yet, he leaves a profound inquiry of the phenomena for further research, calling for more research that makes use of political philosophy in understanding AI and its implications.

Jungherr (2023) introduces a framework that divides the impacts of AI systems on democracy into four levels of abstraction with corresponding areas of impact. Accordingly, AI first affects the individual level by impacts on self-rule. Second, it affects groups of people by eroding equality. Third, it has implications to institutional level through effects on elections. Finally, AI influences the system level through impact on competition between systems, mainly democracy and autocracy. Jungherr thus offers a perspective that he argues is agnostic of democratic ideal theories, rather aiming to "a broader view of the areas where AI might conceivably affect society in ways relevant to the performance and quality of democracy" (Jungherr, 2023, p. 3). Jungherr's perspective is arguably a welcome addition to the discussion of how the impacts of digital technologies on democracy should be approached. Looking at the impacts on separate levels of abstraction can make visible the different layers where those impacts take place, allowing researchers and practitioners to switch between focal points, inspect interconnections and remind us of the multidimensionality of the impacts these systems can have. Nevertheless, I argue it still fails to go deep enough into the foundations of democratic societies, which can be seen in the way in which the impacts brought forth represent only a limited set of aspects and seem to struggle to situate broader phenomena that overlap layers.

Jamie Susskind, on the other hand, takes a broader perspective to social order amid digital disruption and proposes in his book *The Digital Republic* (2022) a new form of republicanism[3], in which regulation is used to

[3] It is worth noting that by republicanism, Susskind refers to "oppos[ing] social structures that enable one group to exercise unaccountable power, also known as domination, over others" (p. 10) rather than any specific party or political orientation.

> "keep the awesome power of digital technology from escaping acceptable bounds of control, and to ensure that tech is not allowed (by design or by accident) to undermine the values of a free and democratic society" (p. 10).

For Susskind, doing so requires following four principles that are put into action through law (p. 11):

1. The law must preserve the basic institutions necessary for a free society,
2. The law should reduce the unaccountable power of those who design and control digital technology, and keep that power to a minimum,
3. The law should ensure that powerful technologies reflect the moral and civic values of the people who live under their power, and
4. The law should restrain government too, and regulation should always be designed in a way that involves as little state intrusion as possible.

Susskind's contribution reveals underlying (and shifting) power relations between technology providers and public government and highlights the importance of regulation in steering technology development towards a more ethical direction. Whereas many proponents of minimalist states also call for as little state intrusion as possible, Susskind establishes a higher threshold for the minimal intrusion that builds on the previous three principles and mainly limits the requirement to ensuring that the mass surveillance of corporation is not exercised by the state (Susskind, 2022, pp. 141–142). In other words, Susskind's idea of what the minimal necessary intrusion is paints a picture of much stronger state than the fourth principle could assume: one where a deliberative democratic state regulates and actively steers technology development towards desirable direction. Indeed, although leaving room for more discussion about the details of this threshold, throughout the book he calls for state intrusion that steers innovation to a beneficial direction and concludes:

> "[W]e should also reject the assumption that the only purpose of economic activity is to generate growth, rather than to produce a society in which life is worth living. The ultimate surrender to market individualism would be to subordinate the scope of our freedoms or the strength of our democracy to the need for economic growth or even technological advance. There is, in occasion, a tension between the logic of capitalist innovation and the public good. And we should not be afraid to say that the public good must sometimes be given priority" (Susskind, 2022, p. 304).

Whereas the contributions of Coeckelbergh, Jungherr and Susskind focus mainly on situating digital technology in existing (or preferable) democratic structures, some

scholars have introduced new forms of democracy that rely heavily on technological developments. For example, Hélène Landemore has proposed a model called *open democracy*, where collective decision-making is facilitated by digital technologies. The democratic ideal Landemore presents falls into the tradition of deliberative democracy, but unlike theorists such as Jürgen Habermas, Landemore rejects the idea of separate tracks for formal and informal deliberation and brings them together on a digital, algorithmically enhanced platform (Landemore, 2021). The task in hand is ambitious, as she argues open democracy to be

> "not just an improved, more participatory or differently representative version of representative democracy but a different paradigm altogether. Its core ideal is to put ordinary citizens at the center of the political system rather than at the periphery, emphasizing accessibility and equality of access to power over mere consent to power and delegation of power to elected elites." (Landemore, 2021, p. 71.)

Landemore thus argues for a paradigm shift in thinking about democratic decision-making and how digital technology can facilitate it. To put her ideal into practice, Landemore proposes a set of mini publics organised on a digital platform, which facilitates political deliberation, leading to a vast expansion of the number of people participating in collective decision-making. Essentially, digital technology would lead to both temporal and spatial opening of democratic deliberation, as it would enable more citizens to participate on deciding upon the same topic, both simultaneously and asynchronously. It would enable sorting and synthesizing of information to support deliberation and ease the information overload citizens face in today's complex information ecosystems (Landemore, 2021).

Yet, looking at the infrastructure that facilitates Landemore's open democracy, it becomes clear that building such platform (which Landemore calls the "Citizenbook") comes with complications, such as ownership questions regarding the platform authority (is the software proprietary or open source and/or distributed?), cybersecurity (how resilient is it against malicious actors?), human autonomy (how much algorithmic filtering of information is acceptable and what becomes censure or manipulation?) and easily spreading mis- and disinformation. Still, Landemore's work is a welcome contribution to for us to understand the different roles pervasive digital systems can have in democratic societies.

What is common in all these conceptualisations of the digital and democratic governance is their tendency to provoke more questions than answers. They remain on a rather high level of abstraction and cover only certain dimensions that constitute a democratic society. They also depend on many "ifs": for instance, for Landemore's democracies to be democratic, the platforms would arguably need to be free from

corporate power over designing the platforms and rather be developed in democratic processes that would grant them the legitimacy of democratic governance. Moreover, for the algorithms to support rather than undermine free formation of political opinions, we would need to mitigate (historical) biases in those algorithms – something that can be considered rather a feature than a bug in the existing systems (for discussion, see, e.g., Simons, 2023, pp. 46–48).

It is understandable that individual articles, book chapters or even short monographs are limited in scope, considering the breadth of the topic in hand. This is also the case of this dissertation. Nonetheless, this implies that we are yet to fully understand the impact of pervasive digital systems on democratic societies, without which maintaining democracies in the age of pervasive digital systems becomes ever more complicated. As a remedy, I propose a perspective that steers attention towards the concept of *democratic resilience,* which has gained attention in the last few years amongst political theorists (Holloway & Manwaring, 2023). I use Merkel and Lührmann's (2021) definition, according to which democratic resilience refers to

> "the ability of a democratic system, its institutions, political actors, and citizens to prevent or react to external and internal challenges, stresses, and assaults through one or more of the three potential reactions: to withstand without changes, to adapt through internal changes, and to recover without losing the democratic character of its regime and its constitutive core institutions, organizations, and processes" (p. 874).

I focus on institutions that form what Rawls calls the basic structure of society (see Chapter 4.1), the task of which is to secure just background conditions for individuals and associations to function in a democratic society. This means studying the resilience of the basic structure when responding to changes provoked by pervasive digital systems. Therefore, the normative ideal for collaboration between institutions in a democratic society comes from Rawls's theory of justice, whereas the conceptualisation of the impact of technological change on contemporary democracies is done through the idea of democratic resilience. This connects Rawls's ideal theory to the reality of contemporary democracies and allows us to distinguish directions and the type of action these changes require. Although this dissertation, too, is limited in scope, this perspective enriches the current picture of the impacts pervasive digital systems on democratic societies as a whole. Doing so, it contributes to the sociotechnical IS research that aims towards ethical and societally sustainable IS development. Rawls's theory and the concept of basic structure is described in more detail in Chapter 4.

Finally, one last remark seems worthy of voicing. It seems that the disruption brought forth by pervasive digital systems has revitalised some of the discussion

about the state of democratic societies (Coeckelbergh, 2024c). For example, the changing power dynamics in data economy have sparked discussion about the relationship between capitalism and democracy (Couldry & Mejias, 2019; Mejias & Couldry, 2024; Susskind, 2022), and the increasing threat of digital manipulation may have made us reflect on not just freedom and human autonomy but what constitutes a good life (see, e.g., Bynum, 2006, 2008; Coeckelbergh, 2022; Stahl et al., 2021; Zuboff, 2019). Perhaps it has made IS researchers and practitioners more aware of the ethical and societal implications of our works. Even if AI did not end up revolutionising the society and human life, it may have paved the way for discussions about what could, for the better. Simons (2023), for instance, does not suggest that changes towards supporting political equality is needed because of AI but because the recent developments have revealed flaws that already exist in the current system.

For IS researchers, this offers a fruitful starting point for interdisciplinary discussions and collaboration that contributes to the strengthening of the sociotechnical dimensions of IS research that has lately been compromised (Sarker et al., 2019), as well as to IS-rooted theorising (Grover & Lyytinen, 2023) that looks beyond organisations towards broader IS ecosystems – a broadening of perspective that has been called for by IS scholars (e.g., Stahl, 2021, 2022). Although the above-discussed literature offers some conceptualisations of this change, I argue that we are only starting to grasp the opportunities provided by technological change for reimagining our existing structures and challenging prevailing beliefs of how societies can and should flourish. This dissertation contributes to this discussion by pushing the field of research towards the third wave of ethics of pervasive digital systems (see Figure 2). Starting with a more technology-focused view to ethics and using AI ethics as an example, I gradually build on the existing knowledge and valuable contributions made in the context of AI ethics and paint the contours of a holistic picture of the impacts pervasive digital systems have on people and democratic societies. The choice of drawing from ethics to explain the impacts of pervasive digital systems on democracy creates solid foundations for this endeavour, as it is a result of millennia of research in understanding what is good or bad, and what constitutes right and wrong. Amidst turbulent technological development, seeking for patterns and interconnections from the common roots in ethics allows for examination of different technologies and combinations thereof, without being trapped into hype cycles. This perspective also enables making observations that are easier to generalise from one technology to another, as it does not focus on ever changing technological features but broader patterns in how pervasive digital systems impact people and societies. I hope this dissertation encourages others, too, in contributing to completing this picture – one piece at time.

# 3 Critical-Political Discourse Studies for Information Systems Research

In this Chapter, I discuss the ontological, epistemological and methodological foundations of this dissertation. Considering that the object of research is complex and multidimensional, transparent description of these foundations is essential. As the goal of this dissertation is both to shed light on ethical and societal implications of AI systems and to provoke action towards a more desirable direction, I rely on constructivist ontology and a critical epistemology. I introduce a research approach titled Critical-Political Discourse Studies (CPDS), which presents a framework for studying discourses around pervasive digital systems that are marked by political dimensions. As one part of this dissertation focuses on the development of the research approach used for the analyses included in this dissertation, a detailed description of the research methodology is presented in Article III. Therefore, in this Chapter, the focus is on the description of the ontological and epistemological choices.

## 3.1 Habermasian constructivism

Despite the increased popularity of approaches drawing from constructivist ontology in IS research, IS researchers rarely discuss their relationship with (social) constructivist ontology explicitly. Whereas a detailed description could take over the scarce space needed for other parts of a research paper – which I admit is the case also of the Articles included in this dissertation – I argue that transparently describing the ontological stance and reflecting on its origins even briefly is crucial for qualitative research. As Jones (1997) notes, only relying on secondary IS sources without seeking understanding of their underlying philosophy can lead to misunderstandings and significant theoretical weakness. He calls for engaging with theorists (such as Foucault and Habermas in case of critical discourse studies), as that is "the necessary price if work in the field is to be taken seriously by others" (p. 108). Ontological assumptions have an impact on all stages of the research, starting from formulating the research question, the choice of methodology and the process

of observing the phenomenon under scrutiny, all the way to the interpretations of research findings. Therefore, they merit a detailed discussion also in this dissertation.

The roots of interpretive and critical paradigms in IS research are tangled with those of social constructivism. In their seminal paper on IS research paradigms, Orlikowski and Baroudi (1991) describe the ontological foundations of the interpretive approach as one where "the social world is produced and reinforced by humans through their action and interaction" (p. 14). Similarly, the social reality of the critical approach "is understood to be produced and reproduced by humans, but also as possessing objective properties which tend to dominate human experience" (p. 19). Considering the seminal nature of Orlikowski and Baroudi's work, one cannot but wonder about the implicit connection to social constructivist ontology that has not been discussed in more detail. Whereas Deetz (1996) was relatively explicit on his connection to the linguistic turn and the discourse perspective, we are witnessing an absence of discussion of the ontological assumptions that led to the increased interest in the role of language in constructing the world and meanings – the foundations for studying the reality and, e.g., the relationship between people, technology and society, through language and action. Orlikowski has in some later works been more explicit about this relationship in some IS research currents, notably in studies where the role of technology is approached through shared interpretations that arise and affect the development of and interaction with that technology – a perspective adopted by some sociologist of technology and information technology researchers (Orlikowski, 1992). This connection has been explicitly stated by, e.g., Pozzebon and Pinsonneault (2005), and Sarkkinen and Karsten (2005). Still, considering the amount of recent literature that draws from sociotechnical theories, the discussion about the roots and the most fundamental assumptions of these IS researchers arising from social constructivism would merit more attention.

The research approach in this paper relies on Jürgen Habermas's version of constructivism, which starts from the assumption that the reality is socially constructed. Habermas calls this reality communicative rationality, which is constructed, or "objectified", in intersubjective communication (Habermas, 1976, p. 8). The roots of Habermas's ontology reach all the way to social constructivism, which emerged as a response to the inapplicability of positivist ontology to research questions that concern non-material social phenomena and cannot be quantified and measured by positivist methods. This constructivist ontology was adopted by many disciplines and fields of research, among which the Frankfurt School theorists are known for their critical epistemology (see the next Chapter). The work of these theorists was taken forward by Habermas, who introduced his theory mainly in the books *Theory of Communicative Action* (English translation in two volumes in 1984 and 1987, originally published as one volume in German in 1981).

Relying on Habermasian constructivist ontology in this dissertation leads to important assumptions. For example, I assume that the way in which we speak about technology in, e.g., regulation and national strategies, is constructive of reality and thus also technology. I assume that we can reveal impacts of ISs on people and societies by studying language about those systems, even if we do not observe the technical construction of a system by decomposing it and isolating the impacts of its different constitutive parts. As shown by Swanson and Ramiller (1997), we shape IS development not just by introducing new ISs but also by attributing meanings to them through speech and action. Conversely, technologies we develop influence the discourse on technology, including regulation, as well as broader phenomena such as what is perceived as a good life or societal progress.

I do not, however, suggest that the adopted approach is always superior to, e.g., positivist perspectives that perceive the world as a physical and social entity independent of humans that can be objectively observed and measured (Orlikowski & Baroudi, 1991). For certain research problems, that is still the most viable approach. Nevertheless, I argue that gaining a deeper understanding of ethical and societal implications of pervasive digital systems requires an approach that emphasises the role of interaction between technology and humans, and how our language and actions influence the way in which those technologies and the meanings attributed to them are being constructed. I argue that ethical and societal implications and the foundations thereof do not have material grounds and thus cannot be "objectively" observed, quantified, and measured in a way that would provide valid and reliable understanding of the current direction we are heading. Rather, we are talking about phenomena that are real but intangible, impossible to reduce into matter the units of which we could quantify in a meaningful way. Therefore, when I study technology as a discourse, I do not study the material of the technology, its features and how they change the material construction of the world and societies. Although technology changes physical infrastructures, too, those changes are not the focus of this dissertation.

I do not consider ethical or societal impacts or change as something we can study even by piecing down the phenomena into events inside the human brain. As Bennett and Hacker (2021) describe:

> "Legal systems consist of laws and not of matter; poems consist of stanzas, not of ink; and revolutions consist of human action and events. The materialist might grant that this is what laws, and poems, and revolutions *consist* of, but deny that they are *made* of anything. We can concede this too. But even if it is true that everything that is made of anything is made of matter, this thesis goes no way to sustain any form of ontological reduction according to which all 'entities' are reducible to material entities. Nor does it support any form of explanatory

> reduction according to which the properties and behaviour of everything that exists are to be explained in terms of the properties and behaviour of its constituent matter." (p. 359.)

Along similar lines, this dissertation focuses on meanings given to and mediated by technology, which result in ethical and societal implications. Such phenomena are not reducible into material entities but are still real and in need researchers' attention. By adding this dimension to the study of technological change we can reach a holistic understanding of the sociotechnical nature of technology and the implications pervasive digital systems have on people and societies.

This choice comes with limitations. Assuming that the object of research is not quantitatively measurable means that the findings of this paper are influenced by many things: they are assumed to be influenced by the researcher's cultural heritage, as well as ideologies and values they have been subjected to throughout their lives. This leaves room for variation in interpretations[4]. This said, to ensure scientific rigour, particular attention needs to be given to critical reflection about the influence of such factors on data analysis and the interpretation of research findings. It requires rooting the study to rigorous justifications that build on ethical and social theories. Else, we risk conducting research that is hardly more than a collection of anecdotes with little to do with scientific research. Fortunately, much has been done to develop tools for IS researchers to structure their efforts stemming from constructivist ontology in a way that yields rigorous research results, contributing to the richness of knowledge on ISs and their role in their context. Next, I describe those structures chosen for this dissertation in more detail.

## 3.2 Critical theory

This dissertation is grounded in critical epistemology. Originating from the Frankfurt School theorists such as Max Horkheimer, Theodor Adorno and Herbert Marcuse, critical theory has gained an increasing attention amongst IS researchers in the past decades and diversified the perspectives to ISs (Cukier et al., 2009; Myers & Klein, 2011; Ngwenyama et al., 2023; Orlikowski & Baroudi, 1991; Stahl, 2007, 2008; Waelen, 2022).

[4] It is also worth noting that despite positivist ontology suggesting a value-free and fully objective approach to research, it also always requires interpretation of research results and what the numerical results of various measurement mean for people and organisations in real-life situations. Consequently, although interpretive research requires more interpretation and is more vulnerable to differences therein, I argue that transparent description of value position and factors impacting the interpretation mitigates many of those issues that are left undisclosed in positivist research.

Critical theories share several fundamental elements. First, they study the socially constructed reality with focus on questions such as freedom, autonomy, and human emancipation (Adorno & Horkheimer, 1979). This often leads to studies that focus on power relations in society (see, e.g., Habermas, 1996; Van Dijk, 2017; Waelen, 2022). Second, critical theories highlight the pragmatic nature of science and knowledge: they aim at not only describing but changing society by challenging existing paradigms and suggesting alternative perspectives (Delanty & Harris, 2021; Habermas, 1976, 1996; Orlikowski & Baroudi, 1991; Stahl, 2008; Waelen, 2022).

Whereas its founding fathers were concerned of the role of technology in society (Delanty & Harris, 2021; Hansen & Caterino, 2019), early adaptations dedicated much less attention to technological advancements, with a few exceptions (notably, Feenberg, 1991, 1999). Since 2000's, however, critical theory is being increasingly used to study contemporary technological phenomena, such as information security (Stahl et al., 2014), AI (Braun & Meacham, 2024; Hollanek, 2023; Hunter, 2024; Krijger, 2022; Waelen, 2022), and digital technologies more broadly (Berry, 2014; Delanty & Harris, 2021). This seems justified, as the key elements of critical epistemology, such as interest in studying power relations, freedom and emancipation, play an important role in the political nature of pervasive digital systems. For example, ISs with manipulatory potential have raised questions around freedom and autonomy, and the increased power of the companies developing pervasive digital systems has sparked discussion regarding power relations in society (see Article V). Moreover, the pragmatic nature of the critical theory compared to other constructivist epistemologies seems particularly appealing in the context of IS research, which has traditionally been tightly linked in IS practice. Therefore, it is here argued to be a viable approach for studying phenomena related to pervasive digital systems, as it enables the researcher to contribute to the knowledge creation, all while providing practical guidance on how the *status quo* can be further improved in the light of the normative-theoretical background.

I have taken as a starting point the perspective of Myers and Klein (2011), who draw from earlier developments in critical IS research (Alvesson & Deetz, 2000; Hirschheim & Klein, 1994; Orlikowski & Baroudi, 1991). They specify six principles for critical IS research (Myers & Klein, 2011, p. 25):

1. Using core concepts from critical social theorists in data collection and analysis
2. Taking a value position that drives the analysis (principles 4-6)
3. Revealing and challenging prevailing beliefs and social practices
4. Encouraging individual emancipation

5. Suggesting improvements in society to overcome unwarranted use of power
6. Improvements in social theories to consider alternative viewpoints and arguments that can further shape the critical theory.

These principles have been successfully used for building new methodologies and theories (Cecez-Kecmanovic et al., 2020; Goede & Taylor, 2019; Kane et al., 2021; Mingers & Standing, 2020; Monson, 2023; Ngwenyama et al., 2023; Sarker et al., 2019). However, their application should not be taken for granted. Despite the seminal nature of the principles in critical IS research (Cecez-Kecmanovic et al., 2020), only a handful of studies have applied them in empirical studies (De Moya & Pallud, 2020; Goede & Boshuizen-van Burken, 2019; McKenna & Chughtai, 2020; Ngwenyama et al., 2023; Rowe et al., 2020; Spil et al., 2021; Young, 2018). Most often the principles are referred to, but the use is not fully reflected in how the empirical study is conducted (e.g., Albertus & Makoza, 2023; Rasmussen & Sahay, 2021; Vaidya, 2019). The approach of Myers and Klein has been criticised for its choice of only Western theorists (Masiero, 2023) and for being "strongly value-laden and partisan, and hence wilfully disruptive" (Clarke, 2020), and lacking attention towards "sensitizing concepts" frequently used by grounded theorists (Charmaz, 2020). Therefore, the principles are in need for some adjustments, as well as a carefully chosen methodology to put them into practice.

First, phenomena around pervasive digital systems are multifaceted and cut across country borders, ethnicities, social groups and communities. Therefore, I call for returning the collective dimension originally suggested by Hirschheim and Klein (1994) – drawing from Alvesson and Willmott (1992) – to the fourth principle: instead of encouraging only individual emancipation, the fourth principle should call for both individual and collective emancipation. In the context of scalable ISs with broad impacts, I consider the collective dimension of emancipation equally important to ensure societally sustainable IS development. The importance of collective efforts in this endeavour has also been underscored by scholars such as Coeckelbergh, who emphasises the need for public deliberation and collective definition of common good when envisioning the direction of AI development (Coeckelbergh, 2024b).

Second, as argued by Hirschheim and Klein (1994), principles need to be sufficiently explicit to facilitate critical evaluation of the object of research. Therefore, in the empirical studies included in this dissertation, i.e., Articles III and IV, I have contextualised the principles and formulated them in a way that makes explicit which theory basis is used in empirical analyses, and how.

Finally, to address concerns of grounded theorists about neglecting contextual, sensitising concepts, I pair the principles with critical discourse studies and political

discourse analysis, which allows for flexibility and sensitivity to aspects arising from the context that an overemphasised focus on the chosen value position – in this study, Rawls's theory of justice – might miss. Description of how that is done in practice is given in the next Chapter, and in even more detail in Article III.

With these changes and additions, the principles of Myers and Klein (2011) provide this dissertation with a background for critical inspection of the impacts of pervasive digital systems. They guide the researcher to bring forth assumptions and contextual elements that affect the outcome of qualitative research so that anyone – the researcher included – can critically evaluate them and suggest alternative, scientifically justified perspectives. The transparency they provide ensures that the proposed alternatives that are to shape further critical IS theorising are not arbitrary but can be challenged in rigorous academic discussion, leading to meaningful developments in the field of IS research. To apply the critical stance in practice, I next outline the methodological choices and shape out the research approach that guides the empirical analyses belonging to this dissertation.

## 3.3 Critical-political discourse studies (CPDS)

Building on critical epistemology, I rely on the tradition of discourse studies (DS) prevalent also in IS research, as well as political discourse analysis (PDA) formerly unknown to the field of IS research. As the nature of the present dissertation is qualitative and the methods less known for the mainstream IS research, I discuss the methodological approach in detail to transparently reveal its fundamental assumptions. I start by describing DS and its critical tradition in IS research and then introduce the PDA method used for empirical analyses. I then explain the research process resulting from the chosen approach.

### 3.3.1 Critical discourse studies: A Habermasian perspective

DS is a family of methodologies that all approach phenomena as *discourses*. DS methodology and methods became established in many fields of study during the linguistic turn of the early 20th century, with three notable branches of origin: French, German and Anglo-American traditions. The latter traditionally focuses on conversations between individuals, which is not the focus of the present dissertation. Therefore, the choice has here been made between the French and German approaches, both of which take a more holistic perspective to the concept of discourse. I build on the German tradition that can be seen to form the basis for critical discourse studies (CDS) also in the field of IS research.

Whereas the French DS tradition draws mainly from formalism and the field of linguistics (with Michel Foucault and Michel Pêcheux as pioneers) (Mazière, 2005),

the current German tradition can be seen to arise from three main sources: hermeneutics, pragmatics and structuralism (Angermüller, 2011). It was thus heavily influenced by the French tradition and notably Michel Foucault, who was in active academic discussion with one of the main contemporary theorists of the German tradition, Jürgen Habermas, until the very end of Foucault's life (see, e.g., Kelly, 1994). Emergence of the critical theory by the Frankfurt School theorists (see previous Chapter) and Jürgen Habermas's application thereof demonstrate the emphasis of the German tradition on pragmatics: language is typically studied through its relationship with society (Angermüller, 2011). Therefore, despite the similarities in the French and German traditions, the latter has still an emphasised focus on the role of language in representation of social norms and power structures, resulting in pragmatic action and social critique.

In IS research, the German tradition has gained a dominant interest over other branches. Increasing attention has been paid to the role of language as a medium of interaction between the IS and its context since the early 21st century, making visible the *semiotic products* such as systems requirements, support calls and user interviews that construct social realities around ISs (Alvarez, 2005). IS researchers have particularly adopted the critical approach to studying discourse: Several IS researchers, Bernd C. Stahl in the forefront (Hur et al., 2019; Stahl, 2007; Wall et al., 2015), have argued for the potential of CDS. On a conceptual level, for example, Wall et al. (2015) argue that CDS can offer ways to fight ideological hegemonies, which can "open debate about the taken-for-granted beliefs and assumptions embedded in academic research" (Wall et al., 2015, p. 259). In the field of IT ethics, Stahl (2007) has applied the CDS approach based on Habermas's theory of communicative action to analyse moral and ethical dimensions of privacy and security and how they relate to ideology (see also Mingers & Walsham, 2010). On an empirical level, in the field of IS development, Auramäki et al. (1992), Alvarez (2002), Pozzebon and Pinsonneault (2005) and Sarkkinen and Karsten (2005) demonstrate how shifting focus from technical aspects towards communication via language can unlock essential knowledge about creating better information systems for human needs. Several authors have suggested approaches that aim to grasp the specificities of discourses mediated by digital technology (see, e.g., Brock, 2018, who introduced a critical technocultural discourse analysis to serve especially critical cultural researchers who study digital media discourses; see also Bouvier & Machin, 2020; Tamássy & Géring, 2022).

Following the choice of ontological perspective (see Chapter 3.1), also the critical stance in this dissertation is based on Jürgen Habermas's critical theory. Habermas's theory of communicative action allows for critical inspection of IS phenomena and how we are constructing digital technologies – whether we are doing so in a way that leads to *communicative action*, which aims at gaining mutual

understanding, or to *strategic action*, which is oriented towards success (Habermas, 1984, p. 286) (see Figure 3). The usefulness of Habermas's CDS has been demonstrated by IS scholars (e.g., Cukier et al., 2009; Mingers & Walsham, 2010; Stahl, 2007).

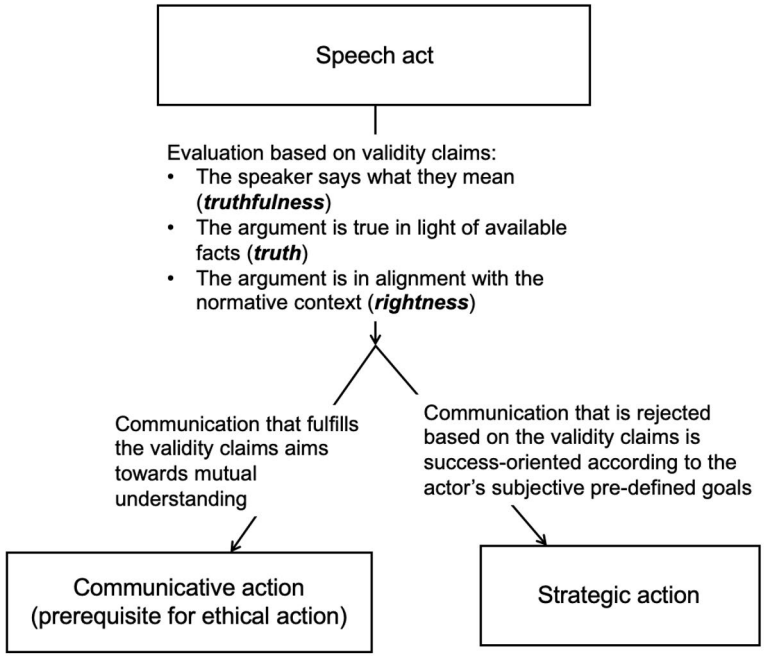


**Figure 3.** Habermas's theory of communicative action defined in a nutshell.

In the context of pervasive digital systems, this perspective is particularly relevant: societal and political concerns around pervasive digital systems come to existence via the use of language and they can either be shared or rejected by others under the validity claims. For instance, whether an argument about an existential threat posed by AI systems is considered misinformation (failing as not true) or disinformation (failing as not truthful) (see, e.g., Westerstrand et al., 2024) impacts its surroundings and what kind of action it promotes. Similarly, regulatory requirements such as the EU AI Act influence the way in which ISs are being developed, how they can be used and how people can seek for corrective actions if ISs do not fulfil the norms outlined in the regulation. Expressing concerns around ISs can lead to either communicative action or to strategic action, depending on the intentions of the agents behind the speech act and how they are received.

The choice of CDS as a methodological approach was informed by the research design, the dedication to the critical epistemology, as well as the continuity of IS research tradition in CDS. Yet, as CDS is not a method but rather a family of methodologies, or a "domain of scholarly practice" (Van Dijk, 2017, p. 2), CDS does not alone give us guidance on how the analysis should be structured and conducted. Therefore, to conduct empirical analyses, the final piece of methodology was added to aid with this, adding to the transparency and replicability of analysis, which I next discuss in more detail.

### 3.3.2 Political discourse analysis

To ensure transparent and repeatable qualitative discourse analyses, I have complemented the CDS with an analysis method called political discourse analysis (PDA) in the empirical studies of Articles III and IV.

I rely on Fairclough and Fairclough's (2013) PDA, which builds on the CDS tradition and thus can be seen as a continuation to Habermas's epistemology. Adding to this background, Fairclough and Fairclough use argumentation theory to study discourses that are marked by a political dimension. For them, PDA is a method for "analysis of political discourse from a critical perspective, which focuses on the reproduction and contestation of political *power* through political discourse" (Fairclough and Fairclough, 2013, p. 17, emphasis by original authors). Unlike many other branches of discourse analysis that build on the critical approach, PDA centres the analysis around argumentation, and more precisely, on the structure of practical arguments. Fairclough and Fairclough see political processes as inherently argumentative. Their PDA combines the descriptive and the prescriptive dimensions of political discourse by both describing the practical arguments and reflecting them against normative context where the argument is presented (Fairclough & Fairclough, 2013, p. 25). In line with the pragmatic nature of the critical approach discussed above, they emphasise *action* that is tied to the discourse and thus its role in changing the society in which the argument takes place (p. 24). The proposed structure of practical argument is illustrated in Figure 4.

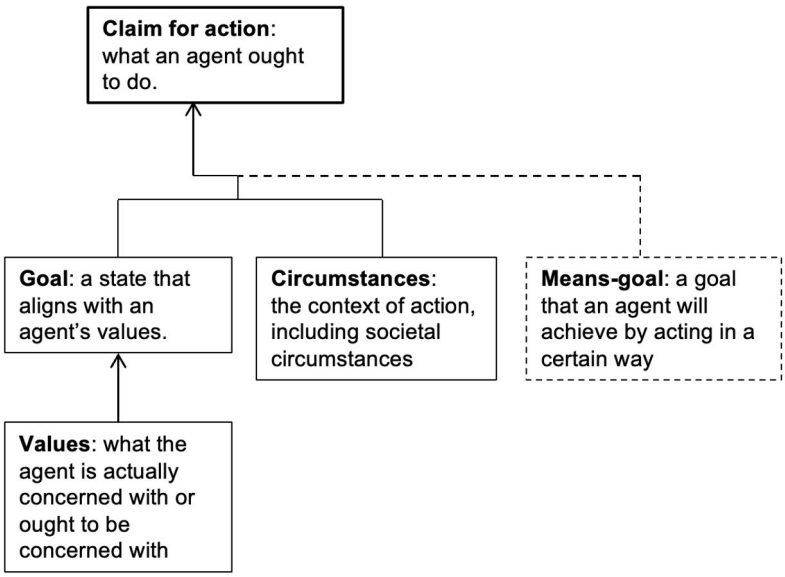


**Figure 4.** The structure of practical argument, adapted from Fairclough & Fairclough 2013, p. 45.

Analysing the structure of practical argument thus starts by identifying elements in the speech act that reflect the values and goals of the agent. In the example of analysing the EU AI Act, the analysis maps the values listed and reflected in the EU AI Act proposals and the final regulation (as demonstrated in Articles III and IV), which are then reflected in the goal setting of the Act that demonstrates the ideal state the regulator wants to achieve with the law. This creates the basis for the claim for action, which is a normative expression of what the agent ought to do. In addition to the agent's goals, the claim for action is influenced by the circumstances, i.e., the context where the argument takes place. Finally, it is possible that there is a means-goal, which is a goal that the agent will achieve through certain action.

In this dissertation, PDA was thus used to guide the empirical analyses around the structure of practical argument. Rawls's theory of justice was used as a reference theory to evaluate the normative premise of the practical argument and the resulting claim for action. In practice, this meant several rounds of analysis, during which the textual data (EU AI Act) was first coded to reveal the structure of practical argument. It was then coded using the key concepts of Rawls's theory of justice, with a particular focus on the principles of justice as fairness. The connections between these two were then analysed and connections made to draw conclusions. The analysis process is described in detail in Article III, with an empirical demonstration in the context of the EU AI Act (for synopsis, see Chapter 5.3).

The choice of PDA to study discourses around pervasive digital systems is motivated by an extensive body of literature that point to the political aspects of digital technologies and the surrounding discourse. This background has been extensively discussed in Article III. It suffices to note here that discourses such as the one around technology regulation is arguably political in nature: it is produced in interaction between different stakeholders from political parties and civil society to businesses and other private organisations. It is a result of intense political negotiations, during which the properties of the technology to-be-regulated is evaluated from the perspective of impacts and their goodness in the given normative framework. In such discourse, the role of technology is not neutral but political (Coeckelbergh, 2022; Susskind, 2022), which I argue is essential to recognise when analysing technology-related discourses. This is not to say that all discourses around technology are political. However, when studying those that are, choosing an analysis method that allows us to grasp the political dimension, make it visible and enable critical evaluation thereof is essential for the rigour of the research.

## 3.4 Discussion: CPDS and its methodological appropriateness in IS research

Together, Habermasian ontology (Chapter 3.1), critical epistemology (Chapter 3.2), and the methodological approach combining critical discourse studies and political discourse analysis (Chapter 3.3) form a research approach called Critical-Political Discourse Studies (CPDS). Conducting research using this approach results in a process that is described in Article III and reproduced in Figure 6. Many advantages and limitations of the choices made while selecting the research approach have already been discussed above, as well as in Article III. However, some remarks remain: is the chosen approach justified in the field of IS research? Is engaging in critical study of ethical and societal implications an endeavour of an IS researcher, or rather of a social or political scientist? Is it problematic that the discourse perspective used here is not endemic to IS research but borrowed?

Firstly, the choice of using existing methodologies as a basis for IS research approach is not obvious. As Grover and Lyytinen (2023) note, we are in need for innovative blue-ocean theorising in IS research, as the recent events in IS development have led to introduction of pervasive digital systems marked by complexity, ubiquity, scalability, general-purpose use and sociotechnical dimensions. They argue that borrowing from neighbouring fields can lead to theories that do not manage to grasp such specificities. However, I argue that turning towards IS-centric theorising might not be the only solution. Rather than choosing “yet another method” -approach, I have chosen an approach that is innovative in a different way: it starts from the research question and dimensions identified in the

object of study, drawing from fields that have known merits in studying such dimensions. I argue that the very pervasive nature of contemporary digital systems requires multidisciplinary inspection, which can enrich the theory basis developed in IS research. Although pervasive digital systems come with new types of impacts, they take place in contexts that have been studied by, e.g., philosophers and social scientists. I argue that this knowledge is valuable when we want to understand the implications of technology on people and societies. In addition, when studying pervasive digital systems, we are studying sociotechnical systems the composition of which depends partly on how we talk about the systems, what kinds of meanings we give them prior to development, how we see their role in society and their purpose of use. Therefore, choosing an approach that focuses on language and how we construct meanings around technology is justified.

Moreover, to avoid falling into a trap of choosing an entirely arbitrary method that fails to grasp the specificities of IS research field, CPDS has strong roots in the IS tradition as it draws from seminal critical IS research and methodologies known for IS researchers. It builds on applications of critical theory presented in seminal IS research papers (e.g., Myers & Klein, 2011). In addition, it brings political discourse analysis to the table as a method, a tool for structuring analysis in a way that takes into consideration elements like those highlighted by, e.g., Grover & Lyytinen (2023). How language constructs technology has been noted already in late 1990s by Swanson and Ramiller (1997), as well as by IS researchers approaching IS phenomena as a discourse (e.g., Cukier et al., 2009; Hur et al., 2019; Mingers & Walsham, 2010; Stahl, 2007; Wall et al., 2015; Westerstrand et al., 2024), which allows me to build on the existing IS research tradition.

In the context of pervasive digital systems, bringing forth a discourse analysis method that is particularly adapted to studying political language is justified by what several scholars have pointed out: pervasive digital systems tend to be political by nature (Coeckelbergh, 2022, 2024c; Susskind, 2022). They have potential to steer human behaviour and decision-making (Coeckelbergh, 2022; Formosa, 2021; Miller, 2021; Muldoon & Raekstad, 2023), which plays a key role in how societies are politically organised. Just because the values are not easily distinguished from the technology does not mean they do not exist (Miller, 2021). Susskind (2022) discusses how companies developing new digital technologies hold an increasing power over our lives and argues that this power should be considered political. It is manifested in how tech companies define the rules in the digital space, exercise surveillance, frame the worldviews of users, and choose the directions of political deliberation (Susskind, 2022). Couldry and Mejias (2019; 2024) formulate this control by denoting that we live in data colonialism, where exploitation of humans through data is considered business-as-usual. AI-enhanced mis-/disinformation that is now close to impossible to distinguish from real has shown to affect political

opinion formation and freedom of election (Alnemr, 2020; Brkan, 2019; Diakopoulos & Johnson, 2021; Kreps & Kriner, 2023; Manheim & Kaplan, 2019; Nemitz, 2018), which threatens democracy (Kilovaty, 2019) – a concern that has been shown valid in the context of parliamentary elections in Europe (Meaker, 2023; Spring, 2024). Whether intentional or not, these technologies impact the quality of information used in opinion formation, giving them a political dimension.

In Article III, I have compared this approach with, e.g., grounded theory methodology, technological frames approach and action research to show how it manages to address several difficulties that persist in its alternatives. Whereas all the mentioned approaches bring valuable contributions to IS research and can be the most suitable approaches to address some research questions, their focus on the organisational perspective limit their applicability to phenomena that exceed organisational borders and collective impacts.

Hence, even if the chosen approach is not necessarily inventive, i.e., creating something from scratch, I argue that it is novel for IS research in the sense described by Grover and Niederman (2021): it alleviates limitations of earlier methods, notably the one pointed out by Grover and Lyytinen (2023) regarding the difficulty of studying pervasive digital systems. It also fulfils the requirements Grove and Lyytinen (2023, p. 53) propose for innovative theories, as it

1. offers a platform for research on unique, identifiable digital phenomena,
2. allows for creation of credible claims supported by available observation patterns that others can build on,
3. creates constellation of coherent, interrelated concepts and logics in a novel manner,
4. offers a proportionate abstraction level to analyse different phenomena,
5. recognises the specificities of pervasive digital systems,
6. supports interventions to manage the impacts trough critical perspective that promotes pragmatic action, and
7. has potential to change the direction and organisation of the discourse on ethical and societal impacts of pervasive digital systems.

This is further demonstrated in the Articles III and IV that use the approach in empirical studies.

Moreover, as noted by March and Smith (1995), IT artifacts and how well they perform is linked to the context where they are introduced, which is why "[i]ncomplete understanding of that environment can result in inappropriately designed artifacts or artifacts that result in undesirable side-effects" (p. 254). CPDS offers a research approach that can build knowledge used as a basis for designing

well-functioning ISs that address real-world problems. Although this approach could be accused of not producing artefacts that can be directly applied by practitioners, it is a step required to use methodologies, such as Design Science Research, that utilise the existing understanding of the sociotechnical environment in designing artefacts. CPDS can thus be seen as a precedent of artefact-oriented research approaches, leading to design informed by ethical and societal impacts of the artefact in question.

Lastly, I share the perspective highlighted by some IS scholars according to which studying ethical and societal implications of ISs is a responsibility of IS researchers. For example, Chiasson et al. (2018) argue that contributing to the direction of development of future ISs can be considered desirable for (if not required of) IS researchers. The constructivist ontology adopted here further highlights this responsibility, as it recognises the role of research in constructing the reality where technologies are built and used. I posit that whether we conduct research with positivist, interpretive or critical lens, we influence the features future technologies have, as well as the societal conditions that steer the development and deployment of those systems in our daily lives. Consequently, taking responsibility in extending IS research to this direction is essential for ensuring future ISs have positive rather than negative impact on people, societies and the environment alike (see, e.g., Walsham, 2012).

In addition to ontology, epistemology and methodology, the choice of a theoretical lens is one of the factors that defines the type of knowledge this research yields. Therefore, keeping this background in mind, I will next discuss John Rawls's theory of justice, how it is applied and why to gain understanding of the ethical and societal implications of pervasive digital systems.

# 4 John Rawls's Theory of Justice for Ethical and Societally Sustainable IS development

To set the ground for what is meant by ethical or societally ideal situation in this dissertation, I draw from John Rawls's theory of justice as fairness. Rawls developed his theory mainly in books *Theory of Justice* (1971, revised in 1999) and *Political Liberalism* (2005). His theory has been praised amongst political, social and moral theorists alike, including his most prevalent critiques. To illustrate, Nozick, who belongs to the latter category, states how political philosophers "now must either work within Rawls' theory or explain why not" (Nozick, 2013, p. 183). Rawls introduced a set of principles of justice that he argues would be agreed upon by free, rational people in the fairest possible setting. In Rawls's theory, however, the subject of justice is not an individual but the basic structure of society, or "the way in which the major social institutions distribute fundamental rights and duties and determine the division of advantages from social cooperation" (Rawls, 1971, p. 7, 1999, p. 6). Hence, institutions belonging to the basic structure of society hold the main responsibility over ensuring fair distribution of primary goods, which is achieved by following his principles of justice.

Rawls drew heavily from Kant when developing his theory, and his theory can be considered a deontological approach due to the way in which it prescribes a set of principles that the institutions have a moral duty to follow. However, it is evident that Rawls's theory is situated in the intersection of moral and political philosophy, as it offers a contractarian theory to how societies should be organised and morally legitimised. This dual nature allows us to use Rawls's theory as a moral guidance for organising our societies in times of pervasive digital systems, recognising the moral duties attributed to us based on our role in the basic structure of society.

Despite the popularity Rawls has gained in several fields of research, one still must start with justifying why Rawls's theory would be an appropriate theoretical framework for IS research. Hence, in this Chapter, I give an overview of the key concepts of Rawls's theory and how they form a normative lens for studying ethical and societal impacts of pervasive digital systems.

## 4.1 Basic structure of society

In Rawls's theory, the main responsibility over securing just background conditions against which the actions of individuals and association take place is attributed to the institutions that belong to the *basic structure of society*. Since we cannot guarantee fully equal conditions in life for everyone, Rawls argues that his principles of justice are primarily targeted to adjust the necessary inequalities distributed by the basic structure institutions. Rawls characterises the basic structure as follows:

> "The basic structure is understood as the way in which the major social institutions fit together into one system, and how they assign fundamental rights and duties and shape the division of advantages that arises through social cooperation. Thus the political constitution, the legally recognized forms of property, and the organization of economy, and the nature of the family, all belong to the basic structure" (Rawls, 2005, p. 258).

The basic structure does not consist of one single actor, such as the state. Rather, it is "an important complex of institutions, given the deep and pervasive nature of its social and psychological effects" (Rawls, 2005, p. 260). Therefore, instead of looking at individual decisions by individual people to deem what is ethical or just, Rawls's defines societal structures that lead to just distribution of necessary inequalities. Rawls justifies his approach by arguing that individual actions can lead to erosion of the background justice even when individuals act entirely fairly. For Rawls, as a result,

> "the invisible hand guides things in the wrong direction and favors oligopolistic configuration of accumulations that succeeds in maintaining unjustified inequalities and restrictions on fair opportunity" (Rawls, 2005, p. 267).

Therefore, when using Rawls's theory to evaluate the ethical and societal impacts of pervasive digital systems, identifying basic structure institutions becomes essential. However, even if reading Rawls word-to-word, defining which institutions belong to the basic structure is not unambiguous. One of the central questions subject to academic debate has been the role of businesses and other private organisations in relation to the basic structure of society: should, for instance, influential companies like Google or Meta be considered basic structure institutions and thus subject to the principles of justice as fairness? Or should the fairness of their actions be reached in other ways, such as regulation? To apply Rawls to present day phenomena, clarification to the composition of the basic structure is of primary importance.

Rawls only makes ambiguous statements when defining the role of private organisations. For example, in *Political Liberalism*, Rawls sets libertarianism in

contradiction with the basic structure. He argues that libertarians consider state like any private association, and that the relationship between citizens and states are seen by them "just like their relation with any private corporation with which they have made an agreement" (Rawls, 2005, p. 264). He continues:

> "By viewing the state as a private association the libertarian doctrine rejects the fundamental ideas of the contract theory, and so quite naturally it has no place for a special theory of justice for the basic structure" (Rawls, 2005, p. 265).

This implies that private corporations do not necessarily fulfil the requirements of institutions belonging to the basic structure. On the other hand, Rawls states that

> "the role of the institutions that belong to the basic structure is to secure just background conditions against which the actions of individuals and associations take place. Unless this structure is appropriately regulated and adjusted, an initially just social process will eventually cease to be just, however free and fair particular transactions may look when viewed by themselves" (Rawls, 2005, p. 266).

That being so, if the role of a private company is such that it partakes in securing just background conditions against which the actions of individuals and associations take place, it is not obvious that it should be left out of the basic structure.

In academic discussion, several interpretations persist. According to scholars who are proponents of the *coercive account*, only institutions with legally coercive power can be seen to belong to the basic structure (Berkey, 2021). Singer (2015), for example, argues that corporations are to be kept separate from the basic structure because that would be in line with Rawls's arguments for freedom of association and individual transactions. Instead, businesses should be regulated from above by the basic structure institutions (Singer, 2015, p. 17). Meanwhile, some consider that the main criterion should be profoundness of impacts an institution has on people's lives (Berkey, 2021). For example, Blanc and Al-Amoudi (2013) argue that the current context of weakening welfare states justifies the revision of which institutions belong to the basic structure. They conclude that private companies should be part of the basic structure because they "bear an effect on the expectations of primary goods associated with relevant (that is, non-voluntary) social positions" (Blanc & Al-Amoudi, 2013, p. 519). They thus base their argument on the purpose of the basic structure given by Rawls (Rawls, 2005, p. 266).

Another and potentially more practical perspective to the debate is offered by Berkey (2021), who argues that the applicability of Rawls's theory to private corporations is not a matter of whether they belong to the basic structure. Rather, the

evaluation should be based on whether they affect people's ability to act as free and equal beings, benefiting from the primary goods (Berkey, 2021, pp. 197–198; 205), which is one of the main goals of Rawls's theory. This perspective does risk excessively bending Rawls's theory, as it contradicts with the original assumption that only the basic structure should be subjected to the principles of justice. It does, however, considerably enhance the applicability of the theory in a setting where power relations are shifting, such as the context of data economy and digital democracies. To enjoy philosophical rigour, however, more development would be needed to justify this position.

I argue that in the context of emerging technologies that have transferred an increasing amount of power over people's lives on corporations (e.g., Coeckelbergh, 2024; Susskind, 2022), excluding all private organisations from the basic structure of society is not justified. Gabriel (2022) argues that in modern societies, the basic structure "is best understood as a composite of sociotechnical systems: that is, systems that are constituted through the interaction of human and technological elements" (p. 220). For Gabriel, AI influences this structure, which subjects its design, development and deployment to the realm of principles of justice as fairness. He justifies this with two reasons: firstly, the societal functions that are relevant in guaranteeing the just background conditions for people are increasingly algorithmically mediated. Secondly, several uses of AI systems have a profound impact on people's lives. Both are integral aspects of Rawls's characterisation of the basic structure of society (Rawls, 1999, p. 6).

Moreover, looking at the argument of the coercive account, according to which only legally coercive institutions belong to the basic structure, we can question whether the coercion needs to be *legally* established in order for it to have a similar impact on societal power relations. For example, if we define coercion as an act of persuading or forcing someone to do something by using threat of punishment, it becomes clear that the power of big technology companies has reached a coercive dimension: if I fail to comply with the rules dictated by, e.g., Meta on its digital platforms, my content will be deleted, or I can be entirely blocked from these platforms. As an increasing volume of public deliberation happens on these proprietary platforms, it would be unreasonable to suggest that the possibility to switch to another product makes this use of power less coercive.

Considering that for Rawls, even his basic liberties are not absolute but can and sometimes even should be modified to fit the "social, economic and technological" context of the society under scrutiny (Rawls, 1999, p. 54), it seems reasonable to do adjustments to our understanding of the basic structure of society when faced by significant changes in such context. Considering the pervasive nature of digital systems (Grover & Lyytinen, 2023), the current technological context reflects a need

for adjustments. In sum, I start from the assumption that some private organisations may belong to the basic structure of society if

- their actions or products have a profound impact on people's lives,
- they limit people's ability to act as free and equal beings,
- they play a key role in securing a just background conditions for people and associations to function in the society, and/or
- they use coercive power over people.

Therefore, providers of digital systems with these characteristics can be analysed through the lens of basic structure of society, which is demonstrated in Article V. Here it suffices to conclude that this dissertation starts from a perspective in which companies developing such systems, e.g., OpenAI, Google, Meta, and many more, should be considered subject to the principles of justice. This does not imply that all providers of AI systems should automatically adopt responsibilities of the basic structure, or that only big companies need to be mindful of these moral duties. Quite the opposite: evaluation of the role of an organisation in Rawlsian society requires reflection of the impacts and their profoundness on people's lives. This might mean that a small company of five people belongs to the basic structure, if they maintain a digital platform used in the gig economy to decide who gets the next job and who does not. I hope this dissertation encourages all organisations from private to public sector to reflect on their role in the basic structure and the following moral duties that could take us closer to fair democracies with positive impact on people's lives.

Before looking at the principles these institutions ought to follow, we first discuss the Rawls's justification for his principles i.e., the original position.

## 4.2 Original position

According to Rawls, his theory concerns *justice as fairness* because his principles of justice are the ones that he argues would have been agreed upon in the fairest possible, universally acceptable setting. Characteristic of a contractarian, he establishes an idea of a hypothetical *original position* to define such conditions. This original position is not a real-life situation but a thought experiment that seeks moral justification for the principles of justice. The main idea of this thought experiment is to show that in the fairest possible circumstances, rational people would choose Rawls's principles over other alternatives – e.g., utilitarianism – to define the fundamentals of a just society.

In Rawls's original position, the principles of justice are defined by people behind a *veil of ignorance*: they know neither their position in the society, nor their personal attributes that might affect the distribution of rights and duties (Rawls,

1999, p. 119). Even one's conception of good is hidden, which leaves people with knowledge only about factors that are relevant for justice (Rawls, 1999, pp. 11; 17).

The parties in the original position are "rational and mutually disinterested" (Rawls, 1999, p. 12), meaning that they do not pay attention to other parties' interests or goals but merely "prefer more primary social goods rather than less" for themselves, thus knowing that factors such as liberties, opportunities and means to promote their aims are worth defending (Rawls, 1999, p. 123). The parties are equal in terms of conception of good and capability of sense of justice (Rawls, 1999, p. 17). For Rawls, these characteristics guarantee that the parties in the original position reflect on the grounds for just society from the place of rationality rather than promotion of the interests attributed to them due to their position in a non-ideal (and in Rawlsian terms unfair) society. The resulting principles of justice defined by Rawls are justified by his arguments on the fairness of the original position. Rawls meant the original position to be such that one can always enter it again (Rawls, 1999, p. 17) and by *reflective equilibrium* adjust the contractarian situation, as well as the principles resulting from it, to eventually end up with principles that are just (Rawls, 1999, p. 18).

It is essential to understand the original position as a justification for the principles and their rationality, which, according to Rawls, leads to adoption of his principles over others (see e.g., Rawls, 1999, p. 13). However, we are not going to enter the original position to define an entirely new set of principles (this is discussed in more detail in Article II). Although it could be justified, as a good fifty years has passed since Rawls originally presented his principles of justice, that is an effort that would merit its own dissertation. As this dissertation is not the one, I contend here with relying on Rawls's principles that would arise in those circumstances, which I next discuss in more detail.

## 4.3 Principles of justice

We have finally arrived at one of the key pieces of Rawls's theory, which is his principles of justice. In this dissertation, these principles have been used to provide a set of AI ethics principles (Article II) and to analyse the Spanish AI strategy (Article I) and the EU AI Act (Articles III and IV) to see how well they align with Rawls's idea of justice. Rawls argues that the principles of justice that would emerge in the original position are the following:

> a) "Each person has an equal right to a fully adequate scheme of equal basic liberties which is compatible with a similar scheme of liberties for all.

> b) Social and economic inequalities are to satisfy two conditions. First, they must be attached to offices and positions open to all under conditions of fair equality of opportunity; and second, they must be to the greatest benefit of the least advantaged members of society" (Rawls, 2005, p. 291).

The first principle (a) is generally known as the basic liberties principle. The second principle (b) is discussed as two separate principles: equality of opportunity and the difference principle. Rawls set his principles in an order of priority, which means that the basic structure institutions must always first fulfil the basic liberties principle before considering the others. Hence, the priority of the first principle means that no basic liberties can be compromised even if it would, e.g., benefit the least advantaged members of society and hence add to fulfilling the difference principle (Rawls, 1999, pp. 53–55). In the context of this dissertation, the order is useful in remedying the issue recognized in existing principles, namely, that they do not offer guidance on how to act in situations when the principles are conflicting (Jobin et al., 2019). With Rawls's principles, navigating conflicting principles is easier, although not uncomplicated, which is demonstrated in depth in Article II.

For the first principle, Rawls offers an incomplete, preliminary list of basic liberties in *A Theory of Justice* and further develops and justifies them in *Political Liberalism*. Accordingly, basic liberties to be equally distributed are:

- Freedom of thought and liberty of conscience
- Political liberties and freedom of association (the right to vote and to hold public office)
- Liberty and integrity of the person (including freedom from psychological oppression and physical assault and dismemberment)
- Liberties covered by the rule of law (Rawls, 1999, p. 53; 2005, p. 291).

Rawls, however, notes that these liberties are not absolute and that they can and should be adjusted according to the "social, economic and technological" context of the society under scrutiny (Rawls, 1999, p. 54). From the perspective of this dissertation, the choice of wording "technological" as a key element that affects the adjusting of basic liberties attracts interest: Rawls does not discuss technological development any further here or in any other place of his works. This is, in fact, one of only three instances where Rawls mentions technology in his works. Yet, it seems that there were some technological changes that he could see as fundamental enough that they could partake in creating a need for adjusting the basic liberties – the very first of his principles of justice. One could argue that the recent developments in technology space could be pointers for us to revisit the set of basic liberties to see whether they manage to grasp the essential even in times of technological

development we are currently witnessing. The analyses in this dissertation, particularly Articles IV and V, give indications that such revisiting indeed seems to be needed. That is, however, an endeavour for further research.

As a response to critique given to him by H.L.A. Hart on the role of basic liberties and their priority, Rawls further elaborates two methods for adjusting the set of basic liberties: 1) a *historical* survey of democratic constitutions to see which ones have traditionally worked well, and 2) an *analytical* consideration based on exhaustion on the set of liberties that are essential for principles agreed upon in the original position (Rawls, 2005, pp. 292–293). The second method means conducting several rounds of iterations to the list of basic liberties, eventually ending up with the fairest according to the criteria of the original position. These methods and possible others have been discussed by several academics over the years (e.g., McLeod & Tanyi, 2021) but not in the context of technology. Here, I contend with stating the need for further research.

The second principle is divided into two sub-principles that apply to the "distribution of income and wealth and to the design of organizations that make use of differences in authority and responsibility" (Rawls, 1999, p. 53). According to Rawls, this must be done so that the positions are open for all (*equality of opportunity*), and that any inequalities benefit the least advantaged (*difference principle*) (Rawls, 1999, p. 53). Out of these two, the difference principle is among the most discussed (e.g., Sen, 2010) and perhaps the most complex elements of Rawls's theory to apply in the current social order. I argue, however, that it is also one of the most powerful tools in his theory, as it steers the attention towards the broader ecosystem where, e.g., pervasive digital systems are being built.

Overall, these principles offer a background for this dissertation to define what is morally acceptable and what is not, steering the focus towards the basic structure of society and how well the institutions therein align their development and use of pervasive digital systems with the principles of justice. Although part of the appeal in Rawls's theory arises from the clarity of the principles themselves and the seeming simplicity of applying them into practical situations, this theoretical choice comes with limitations that need to be considered when evaluating the findings of this dissertation. I discuss these limitations in detail below.

## 4.4 Critique and limitations

Just as any foundational theory subjected to academic scrutiny, Rawls's theory of justice has received critique. Among the most well-known critiques, Robert Nozick responded to Rawls by publishing his own theory, i.e., an entitlement theory of justice (Nozick, 2013[1974]). In *Anarchy, State, and Utopia* originally published in

1974[5], Nozick argues that instead of aiming for an equal distribution of basic goods, everyone should get what they are entitled to, whether through justice in acquisition, transfer or rectification of justice (Nozick, 2013, p. 150–153). Nozick thus proposes a libertarian view to justice, where justice arises as "a product of many individual decisions which the different individuals involved are entitled to make" (p. 130). For Nozick, the process of acquiring holdings is the foundation of justice: "Whatever arises from a just situation by just steps is itself just" (p. 151). This is in contrast with Rawls who argues that we might experience a lack of overall justice even if individual transactions where just (Rawls, 2005, p. 267). Instead, Nozick argues that justice must occur in three instances: the original acquisition of holdings, transfer of holdings, and the rectification of injustice in holdings (Nozick, 2013, pp. 150–154).

Another prevalent critique of Rawls's theory is provided by John C. Harsanyi, who addresses Rawls's logic of arriving to the principles of justice, which is based on the maximin principle: when evaluating what is just, one should look at the worst-case scenario of available alternatives and choose the one with the best worst-case scenario (Rawls, 1999, pp. 132–133). According to Harsanyi (1975), the maximin principle is poorly suited to define social justice, because it does not take into account the likelihood of scenarios to occur and could thus lead to unreasonable situations in cases where highly unlikely worst-case scenarios of otherwise the most just alternative leads to choosing a less just alternative with much more likely – although slightly less bad – worst-case scenario.

Harsanyi's critique is shared by Amartya Sen (2010). In addition to criticising the choice of maximin principle, Sen finds fault with what he calls *transcendental institutionalism*, i.e., the ideal nature of Rawls's theory of justice. He argues that a theory of justice that only provides an ideal situation is not practical. He suggests that we should instead strive towards theories that allow for meaningful comparison of real-life alternatives (Sen, 2010). Moreover, Sen (2010) points out as one of the major difficulties in Rawls's theory that it is poorly applicable to global contexts and global justice. Although Rawls does introduce a continuation to his theory of justice in *The Law of Peoples* that aims to address the need for international collaboration, it only extends to the context of political liberalism, offering "ideals and principles of the *foreign policy* of a reasonably just *liberal* people" (Rawls, 2001, p. 10) (emphasis by the original author). Its main agents are still peoples, each with their "own internal governments" (Rawls, 2001, p. 3). Sen raises a concern about phenomena that are not strictly limited by national borders.

[5] Nozick's theory was thus published three years after the publication of the first edition of Rawls's *Theory of Justice* in 1971. It comes with an extensive discussion about Rawls's theory and its shortcomings.

In the context of pervasive digital systems, Sen's critique seems ever more relevant. Firstly, humans have always been poor at forecasting the future. The complexity of pervasive digital systems challenges our forecasting abilities once again making us uncertain about the impacts of these technologies on people and societies. We are uncertain about the resilience of our societies and democratic institutions when facing technological change (e.g., Coeckelbergh, 2024c). In times marked by an increasing uncertainty and unpredictability, a theory that allows us to evaluate alternative scenarios of the present day (or of short-term future) can seem more feasible than imagining what could eventually lead to an ideally just society in the far future[6]. Secondly, as pervasive digital systems tend to exceed geographical borders, they provoke impacts that would require global coordination. As pervasive digital systems are present in all corners of the globe, including non-democratic governments, ensuring justice of pervasive digital systems can hardly be a concern of collaboration only between democratic governments. Similarly, supranational governance mechanisms typically exceed regional borders. E.g., the EU AI Act is likely to have impacts beyond the EU (i.e., Brussels effect, see Siegmann & Anderljung, 2022), which underscores global nature of digital governance.

It thus needs to be recognised that Rawls's theory is far from flawless and should not be seen as the ultimate theory of justice that surpasses all other alternatives and viewpoints. Correspondingly, it is not my aim to prove that Rawls's theory of justice is the best theory to choose as a basis for ethical IS development. Still, as demonstrated by the praises of his theory even by the above-mentioned critics, Rawls's theory brings in a highly relevant perspective in the discussion of moral foundations of digitalising societies, which we next discuss in more depth.

## 4.5 Opportunities and challenges of Rawls's theory for ethical IS development

I have now described the key elements in Rawls's theory of justice as fairness and its main critiques. Even after addressing the most pressing critics, one could arguably approach ethical implications or pervasive systems just as well from perspectives such as virtue ethics (Bynum, 2006; Constantinescu & Crisp, 2022; Farina et al., 2024; Stahl et al., 2021; Vallor, 2016), utilitarianism (Card & Smith, 2020), or other deontological approaches than the Rawlsian (Aylsworth & Castro, 2024). In recent years, virtue

[6] It is worth noting that it is the sheer distance we currently have from Rawlsian ideal society that provokes this challenge: it can reasonably be expected that reaching Rawls's ideal would require considerable effort and societal change. Therefore, it can be assumed that Rawls's theory is a matter of an ideal currently out of our reach, rather than a potential scenario of short-term future.

ethics seems to have gained popularity amongst IT ethicists studying the impacts AI systems can have on human flourishing (Bynum, 2006; Farina et al., 2024; Kantar & Bynum, 2022; Stahl, 2021b; Stahl et al., 2021; Vallor, 2016), the most recent ones perhaps being motivated by the critique towards the principle-based approach often associated with deontology. Although choosing a Rawlsian approach for this dissertation, I consider these endeavours highly valuable for the accumulation of knowledge about the impacts of ISs on people. As many of the articles included in this dissertation explicitly point out, comparing different ethics perspectives could yield an even more thorough understanding of the phenomena under scrutiny. There are, however, justifications that led me to start my own research path from Rawls.

The first motivation arises from the field of AI ethics, which can be perceived as one of the key fields studying ethical implications of pervasive digital systems. There is currently an overwhelming domination of a principle-based approach in AI ethics, where researchers from different domains propose guidelines and principles to steer AI towards an ethical direction (for reviews, see, e.g., Franzke, 2022; Hagendorff, 2020; Jobin et al., 2019). This reliance on principles has deservedly received critique, notably when applying the principles familiar from bioethics into the context of AI systems (Mittelstadt, 2019) (for further discussion, see Chapter 2). However, it is not the fact of relying on deontic principles that is in the centre of critique but rather their inadequacy to serve as morally justified principles for ethical AI. In fact, many of the proposed principles lack in ethical justifications. As Bleher and Braun (2023) put it, if we do not ground our principles into rigorous ethical reasoning, we end up with principles and operationalisation that risk being "either inappropriate, meaningless, or merely an end in themselves" (p. 10).

Therefore, principles still have a role in how we navigate the complexities of developing and deploying pervasive digital systems. Principles can help us reduce uncertainty and bring uniformity to interpretations in an organisational context. For example, in their commentary, Seger (2022) distinguishes two advantages of ethics principles in the context of AI development. First, inspired by medical ethics, principles provide a useful starting point for articulating rules and requirements for ethical practice. Second, they can cultivate a culture of ethical action by providing principles aligned with cultural norms, rooting policy goals and concrete actions into coherent continuum (Seger, 2022). In short, as several critics point out (Bleher & Braun, 2023; Rességuier & Rodrigues, 2020) principles are not meant to be the end of ethics but rather a tool that – when based on ethics and not mere opinions or ethics washing, if not ethics bashing (Bietti, 2021) – can encourage continuous ethical reflection amongst organisations developing and using digital technologies that have a profound impact on people's lives.

Rawlsian perspective to principles mitigates several issues found in the existing guidelines. First, existing principles rarely address impacts of AI systems on societal

structures and democracy (Hagendorff, 2020), which risks arriving at principles that do not manage to address injustices arising from the institutional setting of modern societies. Rawlsian perspective also alleviates the issue pointed out by Jobin et al. (2019), according to whom existing guidelines do not offer solutions in cases of conflicting principles. In contrast, Rawls's principles of justice are set in an order of priority, with the ultimate priority given to the principle of basic liberties.

Nonetheless, it needs to be noted that the Rawlsian approach is not an off-the-shelf solution that would eradicate the need for contextualising the principles and exercising active ethical reflection while developing and deploying AI systems, which, as e.g. Rességuier and Rodrigues (2020) and Heilinger (2022) note, is an essential part of effective application of ethics in the context of AI. Rawls's basic liberties can conflict with each other, in which cases priorities are not always clear (for more discussion see Article II). Rawls was not, however, oblivious to this issue but indicates that basic liberties can be limited "when social circumstances do not allow the effective establishment of these basic liberties", but only with the one condition: "these restrictions can be granted only to the extent that they are necessary to prepare the way for the time when they are no longer justified" (Rawls, 1999, p. 132). He continues by suggesting that even if such compromises are necessary, the circumstances should still be "sufficiently favorable so that the priority of the first principle points out the most urgent changes and identifies the preferred path to the social state in which all the basic liberties can be fully instituted" (Rawls 1999, p. 132). Therefore, even in cases of conflicting values, we can use Rawls's theory to make justified trade-offs in situations where protection of one liberty seems to come at the expense of another (see Article II). In conclusion, taking a Rawlsian approach contributes to strengthening of the moral philosophical foundations of the principle-based approach, which I demonstrate in Article II.

The second motivation arises from the problem formulation of this dissertation, which is to inspect the ethical *and* societal implications of pervasive digital systems in a democratic context. As Rawls argues his theory to form the most appropriate moral basis for democratic societies (Rawls, 1971, p. viii), his theory combines moral philosophical elements with political philosophy, making it an interesting starting point for a multidimensional research problem. However, despite the academic attention Rawls's theory has gained since its first days, practical applications of Rawls's theory are in general still few and far between (Chandler, 2023, pp. 6–7). The scholars studying emerging technology are often focused on coding Rawls's principles into algorithms (Franke, 2021, 2024; Keeling, 2018; Leben, 2017, 2018), and only few attempts have been made to use Rawls's theory in in broader contexts of digital governance (Douglas, 2015; Gabriel, 2022). I argue that the multidimensionality of Rawls's theory is both its strength and the biggest challenge – it allows the researcher to approach a phenomenon as a holistic entity

with both moral and political dimensions, which reflects the dimensions of emerging technologies (e.g., Coeckelbergh, 2024c; Susskind, 2022). Meanwhile, it challenges the researcher to choose between interpretations and aspects left underdeveloped in Rawls's theory in a way that would benefit from dialogue with other theories, as well as extensive empirical experiments. In the scope of this dissertation, I will not be able to go deep into such comparisons, and the empirical studies conducted need to be complemented in further research. I also argue that this study is a contribution to testing the applicability of Rawls's theory in practice, which can be seen as a valuable effort in theory development in the context of pervasive digital systems, paying homage to efforts in critical theory building in the field of IS research.

One dimension of Rawls's theory that has been left out of the scope of this dissertation but merits a note is the discussion on justice between generations and the resulting just savings principle. According to this principle, society must agree to a savings principle that "insures that each generation receives its due from its predecessors and does its fair share for those to come" (Rawls, 1999, p. 254). For Rawls, this ensures that "any one generation looks out for all" (Rawls, 1999, p. 255), which requires current generations preserve the just conditions for people and associations in society. In the context of an increasing consumption of resources needed to develop and operate complex digital systems such as AI (Bender et al., 2021; Brevini, 2022; Kanungo, 2023; Ren & Wierman, 2024; Rillig et al., 2023), considerations about environmental sustainability of pervasive digital systems through Rawls's lens seems justified. As this dissertation and notably Article V demonstrates, digital systems can also change the fundamental structures of democratic societies, which can have impacts that span over generations. Therefore, discussing justice between generations gains an increasing relevance. However, as it is a dimension that would merit its own article, if not an entire dissertation, it is left out of the scope of the current dissertation. Therefore, I leave the just savings principle as a topic for further research.

To conclude, it is mainly the number of advantages discussed above that led me to choose Rawls's theory of justice as a lens to study ethical and societal impacts of pervasive digital systems. However, in addition, the ambiguities and open questions left by Rawls into his theory make it an interesting starting point for further theory development, as it invites researchers to question the grounds for their interpretations and challenge the assumptions made in a given technological and social context. With the limitations in mind, we now proceed to the section of this dissertation that lays out the findings of the studies that apply the theoretical background to explore how we can best study the impacts of pervasive digital systems, what does the current AI ethics discourse look like around these systems, and what does that mean for the resilience of our democratic societies now and in the future.

# 5 AI Ethics Discourse and its Impacts on Democratic Resilience: Results

It is now time to bring together the findings from the research articles I–V. Below, I give a synopsis of each article, their results and how they contribute to the overall research aims and research questions of this dissertation (Chapters 5.1–5.5). I then give a synthesis of the results and how they respond to the research questions posed in this dissertation (Chapter 5.6).

## 5.1 Article I

Westerstrand, S. (2023). Ethics in the intersection of AI and democracy: The AIDEM Framework. *ECIS Research Papers*, https://aisel.aisnet.org/ecis2023_rp/321/.

Article I is a conference paper presented in the *European Conference on Information Systems* 2023 in Kristiansand, Norway, and was published in ECIS 2023 Research Papers collection. This article presents an analytical framework for approaching the impacts of ISs on democratic societies through ethics. Article I uses AI as a contemporary example of a complex sociotechnical phenomenon the implications of which we need to understand better. The article brings dual contribution to this dissertation: first, it provides an analytical framework that carries through the dissertation in how the relationship between ISs, ethics and democracy is perceived. Second, it offers preliminary empirical findings on the AI ethics discourse and its impacts on democracy in the case of the Spanish AI strategy to validate the analytical framework. In the article, I respond to two research questions:

- Q1: What needs to be studied to address the roots of the relationship between AI and democracy (object of research)?
- Q2: How the object of research needs to be approached to gain better understanding of the phenomenon and enable critique and pragmatic applications (epistemology and methodology)?

In Article I, I argue that we can increase understanding of democratic implications of ISs – AI in particular – by looking at the very roots of the relationship

between the two through ethics. I propose an analytical framework called AIDEM, which is illustrated in Figure 5.

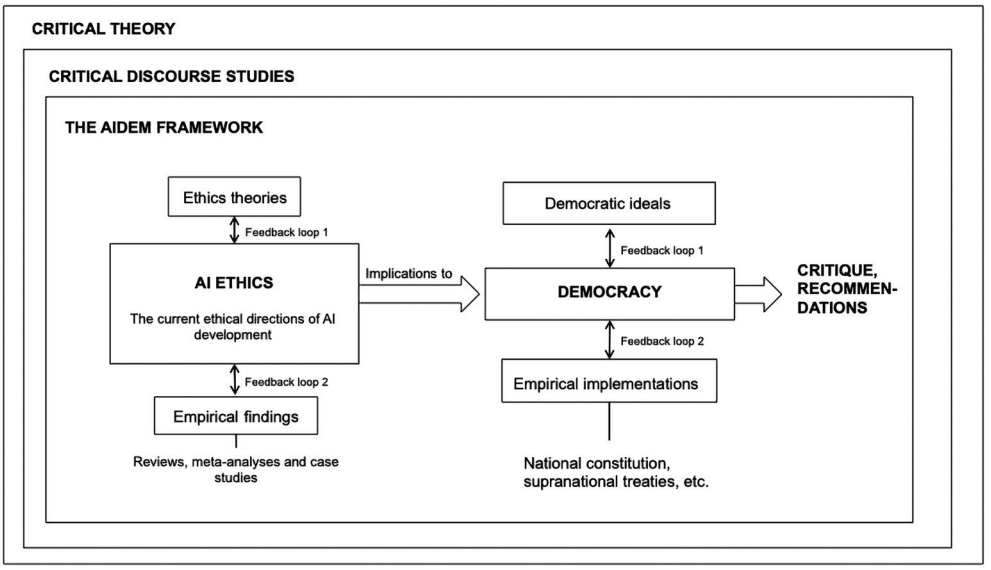


**Figure 5.** AIDEM framework for analysing the impacts of AI on democratic societies, originally published in Article I.

The framework proposes grounding the analysis in critical theory and critical discourse studies. First, the researcher engages in an analysis of the AI ethics discourse to gain understanding of the current ethical dimensions of AI development. These insights are then used to analyse the implications of the revealed ethical dimensions on democracy, which serves as a basis for critique and recommendations on how to direct the development towards a more desirable direction determined by the theory used to define the ethical and the democratic ideals.

In both phases, inspired by the works of Fleuß and Schaal in the field of democratic theory (Fleuß & Schaal, 2019), the researcher engages in two feedback loops – the theoretical and the empirical loop. When analysing the AI ethics discourse, this means engaging with ethics theories to see how the ongoing discourse relates to moral theory and the normative assumptions therein. Then, it requires producing empirical findings that inspect the discourse in real-life context(s). This gives a holistic picture of the AI ethics discourse being constructed in a given context. When analysing democratic implications, this means engaging with both democratic theory and empirical manifestations of real-life democracies to understand what elements of existing democracies AI impacts and how.

To test and validate the framework in practice, the article presents an analysis of the Spanish AI strategy published in 2020, its ethics discourse and its implications on deliberative democracy. Using critical discourse analysis, I analyse the Spanish AI strategy against Rawls's theory of justice as fairness and draw conclusions on the directions it implies for ethical AI development and deployment, as well as implications on deliberative democracy. The choice of the Spanish strategy as research data is explained in Article I.

In terms of theoretical contributions, the analysis implies that the framework is a viable approach for analysing the implications of AI on democracy through ethics. When it comes to the empirical findings, it shows the following findings on the alignment of the Spanish AI strategy with Rawls's principles of justice:

1. The strategy is mostly aligned with the basic liberties but does not give them absolute priority over other strategic goals. Moreover, the Spanish strategy seems to advocate for automation in the public sector without suggesting balancing safeguards against loss of autonomy or democratic legitimacy. The definition of which rights the strategy aims to protect is also vague, which begs a question of its effectiveness in protection of all basic liberties. In addition, no metrics are proposed to measure how well basic liberties are protected over time.
2. Equality of opportunity appears most prevalent in the Spanish strategy, because it highlights the need for support in education and open data initiatives. This support is expressed in several policy measures.
3. Difference principle does not receive support, as there is no indication that the strategy would strive towards AI development that would benefit most the least advantaged members of society.
4. From the perspective of deliberative democracy, if the strategy manages to strengthen basic liberties, it can encourage citizen agency and thus contribute to political deliberation. The emphasis of the strategy on equality and inclusion of multiple perspectives in the development of AI technologies could also have a positive impact on the deliberative dimension of Spanish democracy by increasing opportunities of citizens to participate in collective decision-making between elections. On the other hand, the relative vagueness of the strategy in defining the rights and liberties of citizens could lead to weak implementation in IS practice, which would erode the very foundation of deliberative democracy.

The study also indicated a need for diversification of perspectives, as the analysis only covered one strategy of one country, and only one ethics perspective. It was thus recognised early on that more analyses would be needed to gain a picture of the

broader ethical direction of AI systems development, even if only focusing on the European context.

## 5.2 Article II

Westerstrand, S. (2024). Reconstructing AI Ethics Principles: Rawlsian Ethics of Artificial Intelligence. *Science and Engineering Ethics* 30(46), https://doi.org/10.1007/s11948-024-00507-y.

Article II is a journal paper published in *Science and Engineering Ethics*, with an aim to contribute to the principle-based approach of AI ethics that currently seems to lack in ethical justifications (Bleher & Braun, 2023; Franzke, 2022). Applying Rawls's principles of justice, the article lays out a set of ethics principles for the basic structure of society (inclusive of private organisations) to ensure fair use and development of AI systems. The resulting principles are (Westerstrand 2024, p. 14):

1. Developers and deployers of an AI system must ensure that the AI system does not threaten the basic liberties of any individual.
   1.1. AI systems should not endanger but support the freedom of thought and liberty of conscience.
   1.2. AI systems should not compromise but support political liberties and freedom of association, such as the right to vote and to hold public office.
   1.3. AI systems should not harm but support the liberty and integrity of the person, including freedom from psychological oppression and physical assault and dismemberment.
   1.4. All AI systems should be aligned with the principle of rule of law.

2. The use and development of AI systems should not negatively impact people's opportunities to seek income and wealth. If an AI system is used in distribution of advantageous positions, such as recruitment, performance evaluation, or access to education, it needs to be ensured that
   2.1. the tool is trained with non-biased training data, or appropriate tools are used to mitigate the biases in the final product if no non-biased training data is available (data bias mitigation),
   2.2. the outcome of the use of the tool includes an explanation of the grounds for the outcome it produces (explainability), and
   2.3. the algorithms used shall encourage neither biased results nor the systematic repetition and amplification thereof in, e.g., the feedback loops of a machine learning system (algorithmic bias mitigation).

   If these conditions cannot be met, AI should not be used in the process.

3. All inequalities affected by AI systems, such as acquiring a position of power or accumulation of wealth, must be to the greatest benefit of the least advantaged members of society.

Article II thus responds to the need of ensuring that ethics principles further applied and operationalised by researchers in AI ethics and governance, as well as AI practitioners, are philosophically robust and justified. Looking at the analytical framework presented in Article I (Figure 5), this paper is situated in the theoretical feedback loop, which sheds light on the current direction of AI development from a theoretical perspective. The main goal of the paper is to propose principles that aid institutions belonging to the basic structure of society to reflect on development and use of AI and whether it is in alignment with Rawls's concept of justice as fairness, and if not, why not? What are the underlying justifications for adopting differing principles? What kind of society de their principles envision, if not the Rawlsian ideal? Whereas this paper will not offer answers to these further research questions, I return to some of them later in Article V.

## 5.3 Article III

Westerstrand, S. When Information Systems Go Political: A Research Approach for Political IS Discourse; *Unpublished manuscript, under review in European Journal of Information Systems.*

Article III is a journal paper submitted to the *European Journal of Information Systems* and is, at the time of writing, in revision. It is dedicated to developing a research approach to critical study of IS phenomena underpinned by political dimensions. It lays the methodological foundations for empirical studies in this dissertation, as well as anyone in the IS discipline who wishes to increase understanding of implications of ISs on people and societies. It responds to the need brought forth by several scholars (e.g., Coeckelbergh, 2022; Sarker et al., 2019; Susskind, 2022), according to which we need to pay more attention to the ethical and societal dimensions of emerging digital systems, considering their political nature.

Article III first discusses the motivation for adopting a research approach that addresses political discourse. It points out how emerging digital technologies come with impacts on human autonomy and behaviour (Formosa, 2021; Miller, 2021; Muldoon & Raekstad, 2023) and political opinion-formation and elections (Alnemr, 2020; Brkan, 2019; Kilovaty, 2019; Manheim & Kaplan, 2019; Nemitz, 2018; Spring, 2024). It discusses the shift of power from democratic institutions to private organisations (Susskind, 2022), resulting in forms of data colonialism that exploits humans through collection of data for profit (Couldry & Mejias, 2019; Mejias & Couldry, 2024). To consider the political dimension when studying the impacts of

ISs, the article introduces a research approach called Critical-Political Discourse Studies (CPDS), which aims to remedy the shortcomings of existing approaches in grasping the political dimension of technology. CPDS is based on critical theory and critical discourse studies (CDS). It draws from the principles of critical IS research developed by Myers and Klein (2011) with certain modifications that take the principles one step closer to a concrete level of analysis, ending up with the following principles:

1. Approach the implications of an IS as a discourse exceeding organisational borders
2. Offer a transparent description of the value position and the theoretical background
3. Reveal the rational argumentation structure (underlying values, goals, circumstances, potential means-goals) that lead to a claim for action, relying on the following principles
    a. Reveal and challenge prevailing beliefs and social practices (revelation and challenge)
    b. Encourage individual and collective emancipation (emancipation)
    c. Suggest improvements in society to overcome unwarranted use of power (societal improvement)
    d. Bring forth improvements in social theories to consider alternative viewpoints and arguments that can further shape the critical theory (theoretical contribution)

To add to the rigour of the research approach and ensure repeatability, the approach uses political discourse analysis (PDA) (Fairclough & Fairclough, 2013) as a method for analysis. The resulting research process is illustrated in Figure 6.

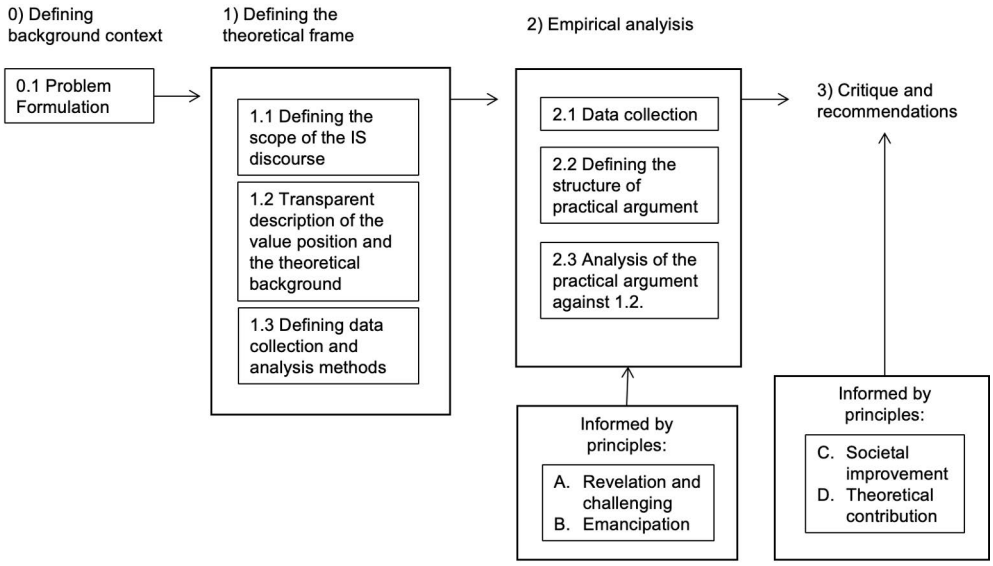


**Figure 6.** CPDS Research process described. Originally presented in Article III.

Details of the different phases, as well as the structure of practical argument used in the step 2.2 to analyse the IS discourse are outlined in detail in Article III. It is tested and demonstrated through an analysis of the EU AI Act, using Rawls's principles of justice as fairness to study what kinds of ethical dimensions does the EU AI Act imply for AI development and deployment. The empirical analysis demonstrates the novelty and usefulness of the approach, which Grover and Niederman (2021) argue to be requirements for an innovative theory.

The framework presented in Article III is not a stand-alone methodology. Rather, it is a research approach that combines elements from critical IS researchers' toolbox in a way that allows for rigorous study of politically loaded IS discourses. The main contribution of this study is to identify an appropriate methodology for studying the implications of pervasive digital systems and to build around it a rigorous research approach that can be used to study various types of political discourses around digital technologies. In addition, Article III offers an empirical contribution in the form of analysing the EU AI Act and outlining considerations for developers, regulators and other stakeholders that play a role in governance of digital systems and their ethical and societal impacts. The role of different actors in this governance is further discussed and developed in Articles IV and V.

## 5.4 Article IV

Westerstrand, S. (2025). Fairness in AI Systems Development: Beyond EU AI Act Compliance. IN Papatheocharous, E., Farshidi, S., Jansen, S., Hyrynsalmi, S. (eds) *Software Business. ICSOB 2024. Lecture Notes in Business Information Processing,* vol 539. Springer, Cham. https://doi.org/10.1007/978-3-031-85849-9_9.

Article IV is a conference paper presented at the 15th International Conference on Software Business (ICSOB), November 2024 in Utrecht, Netherlands and accepted for publication in the conference proceedings. It received a VERSEN Diversity, Inclusion, Equity and Ethics Award as the best paper in the conference theme category. A developed version was invited to the *Information and Software Technology* and is currently in revision. The paper answers the following question:

> RQ: What kind of premise does the EU AI Act lay out for ethical AI development?

Using the CPDS approach, the article is an analysis of the European AI law that entered into force in April 2024 (EU AI Act). The goal of the article is to better understand what providers of AI systems still need to do after compliance in order to reach the Rawlsian ideal of fairness in the context of software business.

To get closer to the everyday work of AI practitioners, the paper focuses on AI systems development perspective, i.e., the providers of AI systems and ethical considerations they would need to address as institutions of basic structure of society. As the EU AI Act is not an ethics guideline but a law resulting from political negotiations, it is reasonable to assume that measures beyond compliance are required from AI providers to ensure ethical AI development. To help academics and practitioners unravel what is already covered by the AI Act and what needs to be considered after compliance, this paper studies the premise that the EU AI Act lays out for ethical AI systems development.

Drawing from critical theory and using John Rawls's theory of justice, the paper shows how the AI Act provides limited support for basic liberties, equality of opportunity and the least advantaged members of society, which calls for attention concerning ethical reflection in the AI system lifecycle to ensure ethically sustainable AI development. The results of the analysis are summarized in Table 2.

**Table 2.** Alignment of the EU AI Act with John Rawls's principles of justice as fairness. Originally presented in Article IV.

| Principle of justice | EU AI Act's alignment with the principle |
|---|---|
| Basic liberties | **Partial alignment**: The goals of the Act are aligned through requirements to follow the EU Charter of fundamental rights and further enforcement for public sector high-risk systems through a requirement for fundamental rights impact assessment (Article 27). However, the enforcement mechanisms notably concerning private sector providers, as well as protection of human autonomy and democratic liberties, are limited. |
| Equality of opportunity | **Partial alignment** through support for SMEs and start-ups and assigning high-risk category to AI systems used in recruitment, promotions, and performance evaluation at work |
| Difference principle | **Weak alignment**, as only minimal protection of rights is offered to the least advantaged members of society, and considering the minimal requirements to GPAI providers. The main beneficiaries of AI development are indicated to be the businesses and their owners. |

Recommendations are given on what kinds of ethical considerations AI providers should include in agile AI development process to strive towards justice as fairness. The recommendations are summarised in Figure 7.

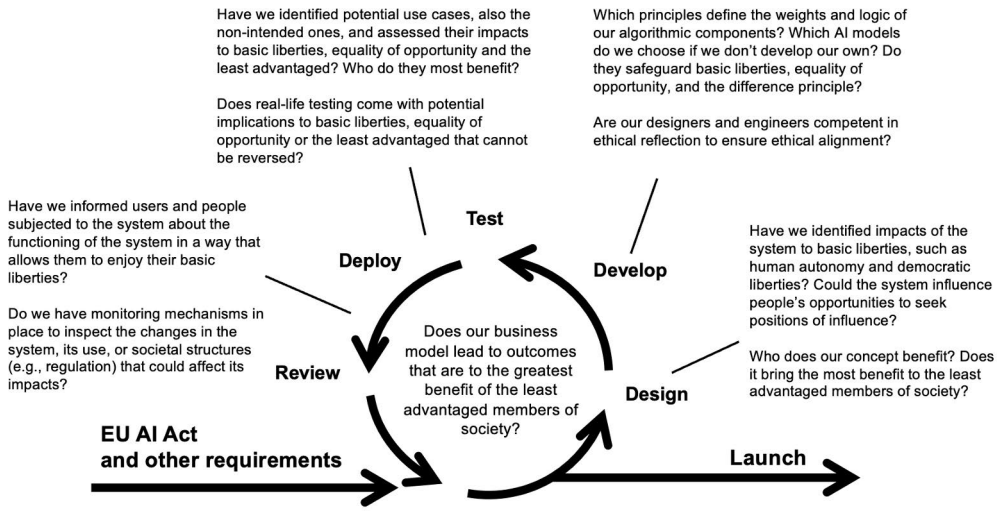


**Figure 7.** Rawlsian considerations in AI development process. Originally presented in Article IV.

Article IV is thus an empirical contribution to increasing the understanding of the state of the AI ethics discourse in IS practice. It is not, however, a complete guide for developers to ethical AI. Laying out task lists and allocating them to different roles in the developer team would be a useful next step to better facilitate the operationalisation of the ethics considerations. It is also worth noting that the AI system lifecycle extends beyond software development and engages many different actors across industries (e.g., Adams, 2021) and requires also societal structures that enable responsible software development in practice (Coeckelbergh, 2024c).

It is also apparent that only asking questions during AI system lifecycle does not yet form a governance framework – this framework is not meant to illustrate all layers of governance identified in previous literature (e.g., Birkstedt et al., 2023). Rather, clarifies what kinds of considerations should be included in AI systems development if we wish to steer it to a more ethical direction. It adds to the existing governance models as it distributes ethical consideration throughout the lifecycle of the AI system, which has shown to be a viable approach in software development contexts (Vakkuri et al., 2021). Whereas regulation can be seen as a top-down normative guidance, ethics functions in this model as a bottom-up consideration that originates from the organization itself rather than an external source of authority. It plays into harnessing the demonstrated potential ethics can have in AI innovation (Bednar & Spiekermann, 2024) when seen as an active process of reflection followed by action (Bleher & Braun, 2023; Heilinger, 2022; Rességuier & Rodrigues, 2020).

This article is thus building our knowledge about the ongoing AI ethics discourse in the context of European AI regulation. Doing so, it commits to the empirical loop of the AI ethics discourse described the AIDEM framework (see Figure 5). It is the final piece in my efforts of mapping the ongoing AI ethics discourse, revealing the ethical dimensions that will be discussed in more detail in Chapter 5.6.

Whereas Rawls's theory and thus the framework resulting from Article IV address questions of social justice, we are yet to address the second dimension of this dissertation: understanding of the implications of pervasive digital systems on democracy, which is the topic of Article V.

## 5.5 Article V

Westerstrand, S. Towards Just Democracies in the Age of Pervasive Digital Systems – A Rawlsian Approach. *Unpublished manuscript, under review in AI & Society*.

Article V is a journal paper submitted to *AI & Society*, in revision at the time of writing. It uses Rawls's theory of justice to seek understanding of the impacts of pervasive digital systems on the fundamentals of democratic societies. It takes the concept of the basic structure of society as a tool for analysing how pervasive digital systems influence the composition of the basic structure and what that implies for

the future of democracy. Article V builds on the existing literature on ethical and societal implications of pervasive digital systems, notably AI ethics.

The analysis starts by analysing whether popularisation of pervasive digital systems change the basic structure of society. It is based on an argument according to which also private organisations, such as businesses, can belong to the basic structure of society if their activities (e.g., products they offer) have a role in securing the just background conditions for individuals and associations to function (Berkey, 2021; Blanc & Al-Amoudi, 2013; see also Article II). The impact of pervasive digital systems on the basic structure is examined through three aspects: the profound impact these systems have on people's lives, the power their providers have over the design and thus the implications of these systems, and the role of regulation as a remedy to managing the impacts and the activities of system providers.

The analysis shows that the profound impact of pervasive digital systems, the power of their providers, as well as the inability of mere regulation to tackle the challenges, imply a need to attribute moral responsibility over social justice also to private organisations as institutions belonging to the basic structure of society. Firstly, pervasive digital systems are being deployed in public governance in processes where decisions have profound impact on people's lives. Simultaneously, conditions for free speech are increasingly moderated by social media companies with the help of algorithms that are used to define to which contents we are exposed and what we can say on their platforms. They impact the conditions for people to earn a living on gig economy platforms. Generative AI tools are useful for people who want to influence others' decision-making and opinion formation, as they enable production of increasingly manipulatory contents (e.g., deep-fakes) that influence our thoughts and actions. It thus seems that the pervasive digital systems influences the just background conditions of contemporary democracies.

Secondly, following the impact of the technologies they produce, organisations designing pervasive digital systems hold an increasing power over people's lives – a form of power that has traditionally belonged to institutions enjoying democratic legitimacy, such as national governments. System providers decide which concept of fairness their algorithmic systems should be based on. Pervasive digital systems are being used for mass surveillance by both governments and tech companies themselves. This shift of power over those profound impacts has led to a situation where organisations such as big technology companies hold power over people's lives without taking the responsibility over ensuring that this use of power is just.

Thirdly, Article V addresses the critique of the coercive account, according to which only organisations with legally coercive power belong to the basic structure of society because they can regulate the businesses. It is discussed how the existing regulation is unable to steer the development of pervasive digital systems in a way that would secure social justice due to the challenges faced by regulatory bodies in

dealing with modern technologies, as well as the heavy lobbying exercised by tech companies to minimise regulatory impact on their actions. Moreover, in AI governance literature, regulation has been shown to be only one dimension of governance measures that are needed for sustainable AI governance.

Therefore, Article V argues that organisations such as AI companies already hold a place in the basic structure but have not yet aligned their actions with Rawlsian conception of justice, which contributes to democratic decline. As a response, the paper proposes a framework for the basic structure organisations towards developing societally sustainable pervasive digital systems, which is illustrated in Figure 8.

Article V thus builds on previous research findings and reveals how the impacts of pervasive digital systems observed in previous research – including the articles of this dissertation – influence the structures of democratic societies. It indicates an urgent need for building democratic resilience in AI development rather than letting the ongoing development lead to an uncontrolled erosion of democracy. Its main contribution to the overall research aims of this dissertation revolves around understanding how the ethical implications of pervasive digital systems also shape societal structures in a way that requires our attention. It adds to the accumulation of knowledge about the impacts of ISs on democratic societies on a fundamental level, which allows us to steer actions of governance towards aspects that could mitigate the negative and encourage the positive impacts. Therefore, Article V is mainly a contribution to the theoretical loop of the AIDEM framework illustrated in Figure 5 but also an empirical one, as it inspects the role of existing institutions in existing democracies and offers a practical way forward. The framework offered as a conclusion aims to build democratic resilience in times of pervasive digital systems. This framework is illustrated in Figure 8.

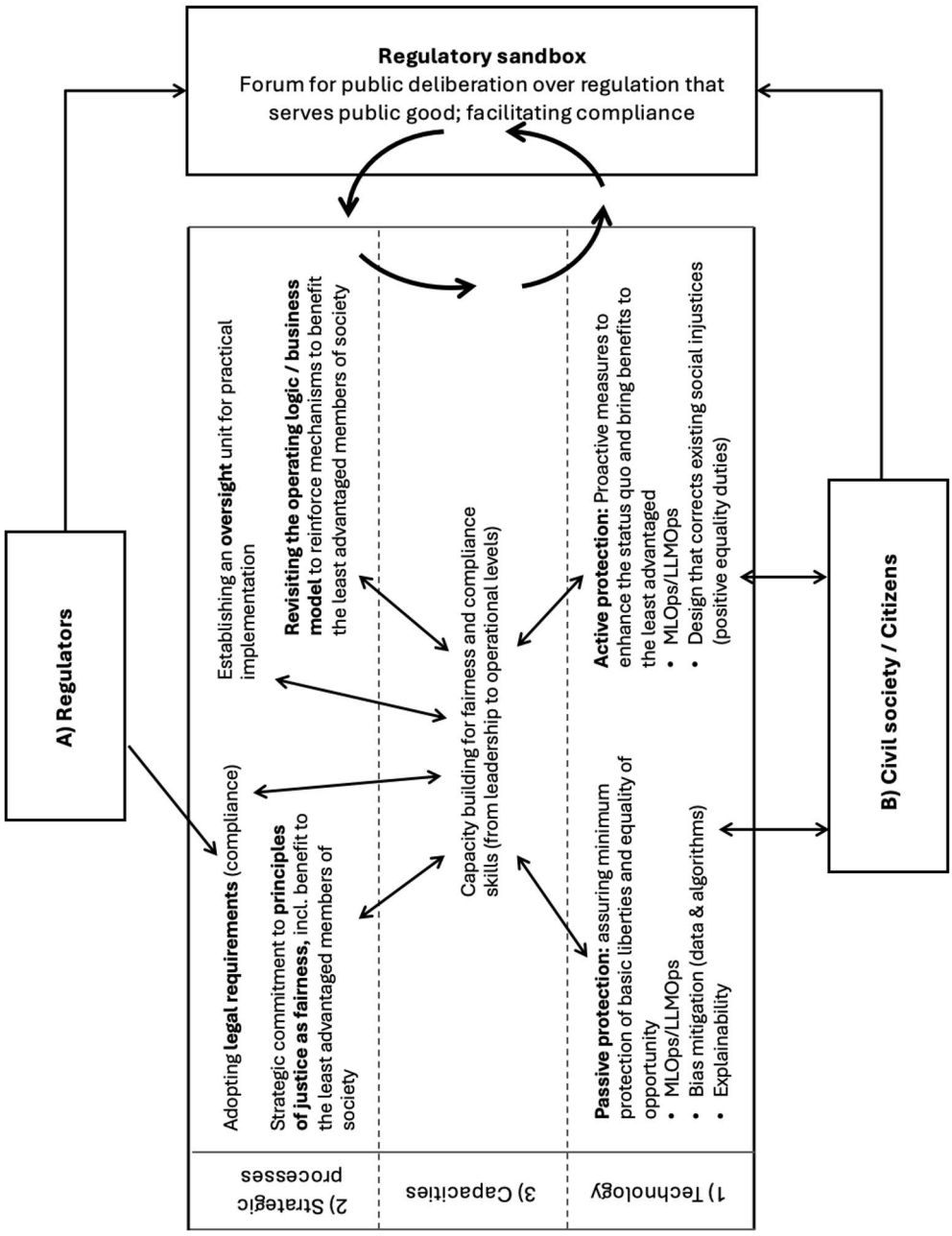


**Figure 8.** Framework for providers for ethical AI development. Originally presented in Article V.

## 5.6 Synthesis of the results

Looking at Articles I–V, the scope of this dissertation is broad. It studies the ethical and societal impacts of pervasive digital systems in a way that aims to both increase the understanding and provide critique and recommendations on how to steer IS development research and practice towards more ethically and societally sustainable direction in a democratic context. It does so by engaging with both theoretical and empirical inquiry. The contribution of individual articles is summarised in Table 3.

**Table 3.** Articles belonging to this dissertation, the corresponding theoretical and empirical contributions.

| Article n° | Title | Primary theoretical contribution | Primary empirical contribution |
|---|---|---|---|
| I | Ethics in the intersection of AI and democracy: the AIDEM framework. | Analytical framework for understanding the ethical dimensions of AI and its implications on democracy | Adding understanding about ethical implications of AI and implications to deliberative democracy in the Spanish AI strategy |
| II | Reconstructing AI ethics principles: Rawlsian Ethics of Artificial Intelligence | Solidifying the ethical and socio-philosophical basis of principle-based AI ethics | Principles for practitioners to guide them in ethical AI development and deployment. |
| III | When Information Systems Go Political: A Research Approach for Political IS Discourse | Research approach for analysing politically loaded IS discourses | Adding understanding about ethical direction of AI in the EU AI Act |
| IV | Fairness in AI Systems Development: EU AI Act Compliance and Beyond | Validation of the CPDS research approach | Adding to understanding of what remains for AI providers to do after EU AI Act compliance to develop ethical AI systems |
| V | Towards Just Democracies in the Age of Pervasive Digital Systems – A Rawlsian Approach | Conceptualisation of the impacts of pervasive digital systems through Rawls's basic structure of society | Adding to understanding of the implications of pervasive digital systems on democracy |

In what follows, I discuss how these results answer the three research questions this dissertation aims to respond.

### 5.6.1 RQ1: Research approach for sociotechnical dimensions of pervasive digital systems

Articles I and III are the main contributions to building a research approach in this dissertation. They respond to the RQ 1: *Methodologically, how could we better*

*increase the understanding of sociotechnical dimensions of pervasive digital systems?* Article I offers an initial analytical framework for approaching the impacts of pervasive systems on democratic societies by showing how ethics can help us understand the impacts of systems such as AI on democracy. With this framework, one can gain better understanding of the impacts of pervasive digital systems on democracies from the perspective of (normative) democratic theory and empirical studies. This enables critique and recommendations informed by scientific research to steer pervasive digital systems towards ethical and societal sustainability.

The Critical-Political Discourse Studies (CPDS) presented in Article III offers a research approach for political IS phenomena. Together, Articles I and III bring clarity to the study of ethical and societal implications of pervasive ISs and open the floor for knowledge creation relevant for both IS researchers and practitioners. This approach is further tested and validated in Article IV, which uses the CPDS to study the ethics discourse around AI systems in the context of AI regulation, proposing critique and recommendations on how the revealed considerations can be integrated into Agile software development by the providers of AI systems.

This demonstrates the innovativeness of the proposed approach called for by Grover and Lyytinen (2023). It provides information that serves both the academic community and industry practitioners that apply the knowledge in their daily processes. It shows how CPDS can offer fresh insights to the ethics discourse around pervasive digital systems to increase our knowledge and understanding of its construction. Moreover, it does so in a way that steers the direction of AI development towards more ethical direction when providers of AI systems adopt the proposed considerations into their practices. Doing so, it contributes to the critical IS research tradition (Mingers & Walsham, 2010; Stahl et al., 2014; Waelen, 2022), strengthens the sociotechnical roots of IS research (Sarker et al., 2019) and adds to the rigour of studying ethics of ISs by rooting the analysis in moral philosophy (Chiasson et al., 2018). The contribution of Articles I and III to the development of critical IS theory also puts into practice the principles of critical IS research (Myers & Klein, 2011), which call for improvement in critical theory and encouragement of social critique, all of which are in the core of the proposed research approach.

### 5.6.2 RQ2: AI ethics discourse

Articles II and IV are the primary empirical contributions of this dissertation in response to the RQ 2*: From the perspective of John Rawls's theory of justice, what kinds of ethical dimensions can we distinguish in the ongoing AI ethics discourse*? I have concentrated on European AI regulation (EU AI Act) and policy, and the discourse around prevailing AI ethics principles, as they are relevant the current AI ethics landscape both in academia and industry (see Chapter 2).

Article II is a conceptual paper that takes the Rawlsian perspective to principle-based AI ethics. It discusses contemporary phenomena that clash or align with Rawls's principles of justice and proposes a set of principles for organisations to aim towards ethical and societally sustainable AI development. Article IV, on the other hand, is an empirical study on the premise of the EU AI Act for ethical AI development, revealing considerations that remain for AI providers to do after compliance if they wish to aim towards Rawlsian ideal of justice as fairness. Article IV thus focuses on software businesses and other industry actors developing AI systems, increasing the practical relevance of the research results produced in this dissertation. In addition, the empirical studies accompanying the theoretical developments in Articles I and III add to the results: Article I includes an analysis of the Spanish AI strategy and Article III an analysis of the EU AI Act, both in the light of Rawls's theory of justice. In contrast with Article IV, the analysis in Article III takes a broader perspective of organisations developing and deploying AI systems, whereas Article IV focuses particularly on AI systems providers.

As a synthesis of these results and the associated background literature, we can distinguish four sub-discourses that characterise the ongoing AI ethics discourse in the context of AI regulation and policy. These sub-discourses are described below.

1. **Moral duty to follow principles, guidelines and regulations with weak ethical foundations**. All the articles in this dissertation recognise the dominance of the principle-based approach in ethics and its limitations. Although many of the principles proposed around the globe lack ethical justifications (Franzke, 2022; Jobin et al., 2019), and those with justifications rarely take an explicitly deontological stance, they call for a moral duty to steer technology towards an ethical direction. For example, one of the most cited literature reviews of AI ethics principles conducted by Jobin et al. (2019) reveals a plethora of principles that have been introduced in both public and private sectors without being legally binding. This emphasis on principles was examined notably in Article II. These limits include weak connections to ethical theories (Franzke, 2022; Heilinger, 2022; Jobin et al., 2019), lack of discussion on impacts on democracy and societal structures (Heilinger, 2022) and difficulties with

prioritising between principles when they conflict (Jobin et al., 2019). The set of Rawlsian principles for fair AI proposed in Article II aim to remedy these challenges by strengthening the moral-philosophical foundations of the ongoing AI ethics discourse.

2. **Protecting individuals from infringement of basic liberties**. Empirical studies in Articles I, III and IV revealed an emphasis on the calls to protect individuals' basic liberties. When reflecting the ethics discourse in the EU AI Act and the Spanish AI strategy against Rawls's principles of justice, both the goal setting and the proposed actions provided strongest support for the basic liberties out of all Rawls's principles. Despite lacking in prioritisation and mainly consisting of protection against infringements rather than active promotion of the realisation of the basic liberties, there seems to be a consensus in the European context around the need for special protection of basic liberties in the context of AI development.
3. **Regulation alone does not guarantee ethical AI**. Articles I, III, IV and V also indicate that the current regulatory landscape around AI is alone insufficient in guaranteeing ethical AI development in the Rawlsian ideal. Rather, additional measures are needed in strategy, governance and AI development practices to steer the AI discourse towards ethical direction, mainly due to uncertainty in efforts aimed at protecting the equality of opportunity, and a clear lack of aspiration towards generating most benefits to the least advantaged members of society.
4. **Call for governance and technical operationalisation**. Finally, all articles reflected the calls for operationalising ethics principles and other requirements into pragmatic actions along the system lifecycle. The findings of the Articles confirm arguments of previous research that call for translating ethics into Machine Learning Operations (MLOps) task lists and organisational processes to make it easier for organisations to follow them, as mere principles and regulation can be abstract and insufficient in ensuring ethical alignment (Birkstedt et al., 2023; Ibáñez & Olmeda, 2022; Mäntymäki, et al., 2023; Morley et al., 2021, 2023).

The revealed construction for the AI ethics discourse in the given context is illustrated in Figure 9.

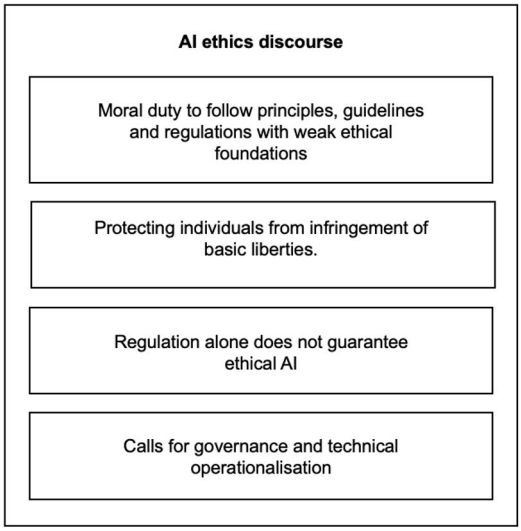


**Figure 9.** AI ethics discourse as revealed in the present dissertation and its constituent sub-discourses.

In sum, studying the language used in AI policy and regulation reveals a will to promote ethical AI development. Meanwhile, the sub-discourses indicate a power imbalance between actors that express this will, and actors who hold the power to act on that will. Whereas theory, academic discourse and regulation call for ethics and protection of individuals, and collectives of people, organisations such as the providers of AI systems are yet to act. The results of this dissertation thus beg the question on the motivations of different actors in steering AI development. In Habermasian terms, we could ask: are stakeholders that take part in, e.g., drafting the EU AI Act, advancing their own interests in a strategic game, or do they transparently present their rational arguments that contribute to communicative action? Do the arguments fulfil the validity claims of sincerity, truth and rightness? It seems that the validity claims are not always fully met (Westerstrand et al., 2024). Considering that the results of this dissertation show an active discourse around AI ethics yet little indication of ethics being an active part of AI development practice, it seems that there is still room for public deliberation on the roles of different actors in ensuring ethical AI development.

### 5.6.3 RQ3: Implications of AI on democratic resilience

Finally, we have arrived at the RQ 3: *In the light of the response to RQ2, what kind of implications do pervasive digital systems have on democratic resilience?* The main contribution to answering this question is given in Article V, which draws from the observations constructing the AI ethics discourse and studies their implications

on Rawlsian basic structure of society. Article V highlights the need for attributing moral duty over social justice also to providers of pervasive digital systems, such as big tech companies, as they play a key role in realising social justice in times marked by rapid development of pervasive digital systems. Article V shows that these systems have influenced the constitution of the basic structure but the institutions new to the basic structure have not yet aligned their actions with the principles of justice. Looking at the three potential reactions of democratic societies to express democratic resilience outlined by Merkel and Lührmann (2021) – "to withstand without changes, to adapt through internal changes, and to recover without losing the democratic character of its regime and its constitutive core institutions, organizations, and processes" (p. 874) – we are yet to see signs of the basic structure being able to adapt and to recover. Rather, the findings of this dissertation imply a risk of losing the democratic character of the current regimes due to the shift of power over life-changing decisions and infrastructure of democratic deliberation from democratic institutions to providers pervasive digital systems.

This implies an urgency in responding to these changes in democratic societies, calling for action to ensure democratic societies are built on just foundations. For many democratic theorists, including Rawls, democracy and justice are inseparable (Buchanan, 2002; Cohen, 2003; Pettit, 2012), and thus eroding the just background conditions can be seen to have an eroding rather than strengthening effect to democracy. In times when the state of democracy is characterised with a word as strong as *crisis* (Przeworski, 2019), and concerns around democratic backsliding seem to be persistent (Bermeo, 2016; Grillo et al., 2024; Wolkenstein, 2023), such course of development seems to contribute to the convoluted network of events that has raised concerns about the survival of democratic regimes. This puts democratic resilience to a challenging stress test: as the changes are occurring at a dimension as fundamental as the basic structure of society, we cannot but wonder whether contemporary democracies still have enough resilience left to be able to recover and improve.

In addition, Article I offers one – although succinct – perspective to the impacts of the AI ethics discourse on deliberative democracy. It shows how the emphasis on basic liberties can promote emancipation and citizen agency, as well as promotion of public deliberation that is essential for deliberative democracy. Similar positive effects also arise from the goal of increasing inclusion of people from various perspectives and backgrounds into the discussion. However, as the rights to be protected, as well as the actions to fulfil the goals remain vague, there is little guidance on how organisations such as providers and deployers of AI systems can ensure their systems promote democratic resilience rather than erode it.

To conclude, the findings of this dissertation show that the current ethical direction comes with a risk of eroding democratic resilience due to the ambiguities

in responsibilities over securing just background conditions for people and associations to function in society, as well as weak political support for political liberties that would be required for citizens to meaningfully contribute to the collective decision-making. For Rawlsian ideal of democracy, we are thus witnessing a direction towards weaker democratic resilience, which calls for action to steer the AI ethics discourse and the resulting democratic impacts towards a more ethical and societally sustainable direction.

### 5.6.4 Critique and recommendations: artefacts

The analysis of the findings of this dissertation have thus far been more interpretive than critical. In the spirit of critical IS research, the Articles included in this dissertation have given normative guidance on how the current course could be steered towards a more ethical direction. The components of such a model consist of frameworks developed to both researchers and practitioners in a way that recognise the role of different stakeholders in ethical digital ecosystems. The focus is given to institutions that belong to the Rawlsian basic structure of society. The artefacts contributing to this critique are listed in Table 4.

**Table 4.** Artefacts presented in the Articles of this dissertation and their primary target institution.

| Article n° | Title | Artefact | Primary target institution |
|---|---|---|---|
| I | Ethics in the intersection of AI and democracy: the AIDEM framework. | AIDEM: Analytical framework for studying impacts of AI on democracy | Researchers |
| II | Reconstructing AI ethics principles: Rawlsian Ethics of Artificial Intelligence | Ethics Principles for Fair AI | Providers and deployers of AI systems |
| III | When Information Systems Go Political: A Research Approach for Political IS Discourse | CPDS: Research approach for studying ethical and societal impacts of political information systems discourses | Researchers |
| IV | Fairness in AI Systems Development: EU AI Act Compliance and Beyond | Agile framework for ethical considerations in AI systems development | Providers of AI systems |
| V | Towards Just Democracies in the Age of Pervasive Digital Systems – A Rawlsian Approach | Framework for Rawlsian governance of pervasive digital systems | Organisations belonging to the basic structure of society |

First, the analytical AIDEM framework presented in Article I and the CPDS research approach is provided in Article III are tools primarily for researchers to seek for understanding of ethical and societal implications of pervasive digital systems. Whereas the CPDS approach is clearly a research approach, the AIDEM framework can also be used by industry actors and regulators to conceptualise the impact of their actions on people and societies. Second, to practitioners, this dissertation offers a set of ethics principles for organisations that belong to the basic structure of society and provide AI systems (Article II). It also demonstrates how the moral duty to follow these principles translate into considerations in AI businesses (Article IV), and which stakeholders play a key role in the basic structure to that secure the just background conditions for people and associations to function in society (Article V).

Together, the proposed actions developed in dissertation form an initial high-level model of the basic structure of society in the context of the development of pervasive digital systems that gives normative guidance on how the different institutions can steer the development of these systems to a more ethical direction that strengthens democratic resilience. This model is illustrated in Figure 10.

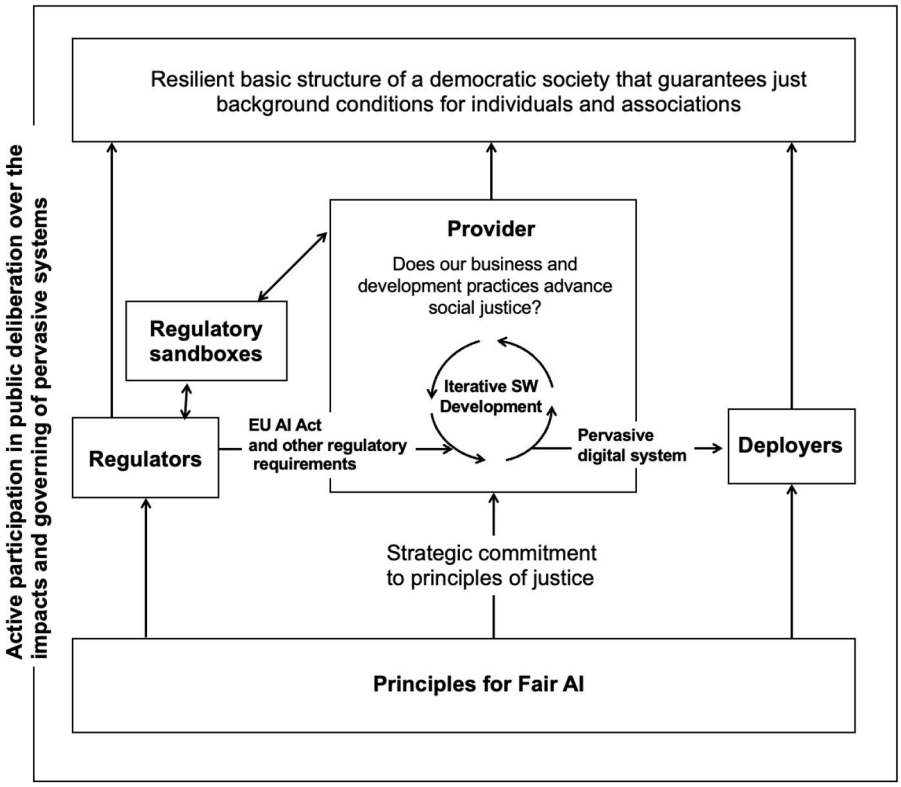


**Figure 10.** A model for a resilient basic structure of society that guarantees just background conditions for individuals and associations to function in the context of pervasive digital systems.

The model represents the role of key institutions in a resilient basic structure of society in the context of development and deployment of pervasive digital systems. At the top of the model, the aim is formulated to fit the Rawlsian ideal of fair society, where the basic structure of society guarantee just background conditions for individuals and associations to function in society.

At the bottom of Figure 10, the principles for fair AI (see Article II) are proposed as an ethical foundation for institutions belonging to the basic structure of society and a) regulating, b) providing, or c) deploying pervasive digital systems with AI. Pervasive digital systems often include an AI component, which is why it is reasonable that the principles are in this context particularly directed to AI systems. This, however, needs future research to explore how well the principles apply to pervasive digital systems without AI components, and what adjustments might be needed.

The focus of this dissertation has been on *providers* of pervasive digital systems, which is reflected in the model and the level of detail it gives to providers in comparison with regulators and deployers. For regulators, the model underscores their duty to regulate through both hard and soft law in a way that advances justice as fairness. This means providing requirements for the providers that contribute to their efforts in fair development of digital systems. For providers, the model calls for strategic commitment to the principles of justice, as well as integration of Rawlsian consideration into software (SW) development practices. Assuming that the practices and the strategic commitment already prioritise the basic liberties, and considering the strong regulatory support for basic liberties, the focus is here on whether the business model of the provider brings the most benefits to the least advantaged members of society. For more detailed description of the considerations in iterative development, see Article IV. To balance between regulation and innovation, regulatory sandboxes are suggested as a forum for regulators and providers to deliberate on how social justice could be best realised (see Article V). The systems developed according to this model are then used by the deployers, who should align their use of the system with the principles of fair AI.

Finally, all actors should engage in public deliberation on governing pervasive digital systems to strengthen the democratic legitimacy of the use of power and attributing citizens agency as key stakeholders in adjusting the model and the following practices further – contributing to the Rawlsian ideal of democracy and civic engagement. Establishing continuous practices for doing so is thus needed. However, democratic participation and civic engagement are topics that have received little attention in the present dissertation (mainly Article V), making these findings are a preliminary attempt to show the relevance of participation for democratic resilience. A more developed idea of the design of such practices and their moral and political philosophical justifications is regrettably beyond the scope of the present dissertation, and a much-welcomed topic for further research.

# 6 Discussion

> "Instead of depending solely on social relationships, digital pervasive phenomena intertwine the social and material in unprecedented ways. Now technology constitutes an inherent part of organizational phenomena and the social cannot be separated from its digital mediation" (Grover & Lyytinen, 2023, p. 48).

It is time to discuss the results in relation with the goal of this dissertation, which is to shed light on ethical dimensions of pervasive digital systems and the implications those dimensions have on democratic societies, as well as to provide critique and recommendations for action. The research that has contributed to this dissertation has been conducted mainly between years 2021 and 2024, during which a lot has happened in the practice of developing and deploying pervasive digital systems, as well as IS research studying the implications thereof. Along with the changes in the field, my own thinking has developed over time, leading to both successes and challenges in the overall research project spanning over four years. It is thus now time to discuss what the findings synthesised in Chapter 5 mean for ethical development of pervasive digital systems and democratic resilience.

## 6.1 AI ethics discourse: What is our direction and are we happy about it?

To start with, this dissertation has succeeded in increasing our understanding of the current direction where the practices in developing and using pervasive digital systems are taking us. I have revealed how the current AI ethics discourse highlights principles but lacks in depth when it comes to moral philosophical justifications. I have identified a call for operationalisation of principles in AI development practice but also noted how it fails to grasp the essence of ethics that would allow practitioners to do so sustainably. The lack of ethics in AI ethics was acknowledged early on, which built the motivation and background for this dissertation and paved the way for the choice of Rawls's theory of justice as a theoretical background (see Chapters 2 and 4). The articles in this dissertation brought forth the challenges of misunderstanding ethics even more, as it was shown how regulation alone fails to

respond to the needs of ethical AI development and how some organisations belonging to the basic structure – i.e., providers of pervasive digital systems such as AI – have not committed to securing social justice.

From the perspective of ethics of pervasive digital systems, these tendencies are concerning. If we reduce ethics into mere principles and focus on technical operationalisation of those principles into technical features and development tasks, we increase the distance between actions and their moral justifications. When a developer then makes design decisions, chooses between AI models and decides upon data preparing methods to deploy, the processes that guide them do little in equipping them with skills for ethical reflection. What should the developer consider when the available options come with difficult trade-offs between values and principles? Are designers ready to assess the implications of their concept on a societal level, considering the least advantaged and how the solution impacts their situation? Bleher and Braun (2023) discuss the challenges in different methods for operationalisation. They note how approaches that aim to embed ethics into practical processes by involving ethicists as experts, as well as principle-based approaches that rely on pre-defined principles, fail to reflect on the ethical justifications of what they call ethical action. They also note the challenge of relying on principles that have been shown in, e.g., literature reviews to be empirically relevant, as their relevance and interpretations in different context might vary considerably (p. 10).

If we want to ensure that development and deployment of pervasive digital systems is guided by ethics and theoretically robust ethical justifications, we need to promote ethical reflection amongst relevant actors. From a Rawlsian perspective, focus should be given primarily to the basic structure of society, as it has power and thus also the moral duty to shape societal structure towards justice as fairness. In this dissertation, principles are provided to help with high-level steering and trade-offs (Article II), guidance is given for providers of AI systems on what types of consideration should be included into the AI systems development lifecycle (Article IV), and a framework is developed for institutions belonging to the basic structure of society to facilitate the adoption of their role as those securing the just background conditions in society (Article V). These actions are illustrated in Figure 10 above.

I argue that choosing Rawls's theory of justice has been the key factor enabling a meaningful contribution to the conceptualisation of the status quo, as well as providing critique and remedies to the current AI ethics discourse. One such contribution of adopting Rawlsian perspective revolves around the complexities of fairness in machine learning. Whereas it would be tempting to think that simply removing attributes related to sensitive characteristics (e.g., age, gender or ethnicity) from algorithms would mean those attributes do not impact the final decision, we know that proxies and redundant encodings still often lead to harmful discrimination (e.g., Hacker, 2018). As discussed by Simons (2023), when such algorithms are

presented and accepted as neutral, it becomes even harder for us to trace back the grounds for discrimination (Simons, 2023, pp. 63–64). Experience from the US shows a tendency to interpret discrimination law in a way that favours anticlassification principle (the way of justifying non-discrimination by e.g., removing protected attributes from an algorithm) over antisubordination principle, the latter of which would call for action towards eliminating systemic exercise of power that maintains inequalities in the first place (Simons, 2023, p. 73). Therefore, a disproportionate emphasis on non-discrimination through anticlassification in the context of pervasive digital systems can in fact lead to harmful discrimination that is harder to detect and to prevent, as the discriminatory algorithms can be presented as just due to the lack of attributes that would explicitly give away discrimination against protected groups.

In contrast, Rawls's principles go beyond non-discrimination. Firstly, aligning digitalisation with pre-identified basic liberties would encourage the providers of pervasive digital systems to demonstrate how the patterns coded into their ISs explicitly support the basic liberties and the equality of opportunity principle. Secondly, the difference principle makes visible discrimination that can be considered fair, e.g., positive discrimination to the benefit of the least advantaged members of society. Programming an algorithm explicitly to improve the situation of the least advantaged could make the efforts towards justice as fairness easier to demonstrate. In other words, Rawls's principles do not merely call for absence of injustice but presence of justice and fairness, which requires action from the basic structure of society. Adopting the interpretation used in this dissertation according to which providers of AI systems belong to this basic structure, the accountability over demonstrating such proactiveness would be attributed also to the providers of pervasive digital systems. Rather than programming in a feedback loop that maintains or encourages inequalities (e.g., an algorithm that targets women with lower income positions and men with higher income positions), algorithms could be used to create positive feedback loops with corrective impacts on individuals and groups of people (e.g., recommending women positions with income equal to that of positions recommended to men). What these are exactly should be subject to public reason.

It thus seems that the Rawlsian perspective helps make visible the value positions taken in the name of justice as fairness. It offers better means to evaluate whether pervasive ISs are ethical or not and helps us identify who is responsible. This presumably only works if we accept Rawls's justification for why his principles form an adequate basis for fair society. This indeed seems to be the type of discussion that we currently lack on a societal level: whose ideal of justice are we pursuing? Why do we believe that principles such as transparency, justice, non-maleficence, accountability, privacy, beneficence, freedom and autonomy, trust, dignity,

sustainability, or solidarity are worth special attention when developing ISs (as indicated by, e.g., the much-cited literature review by Jobin et al., 2019)? Where will they lead us – towards human flourishing, increased utility, or something else? For whom?

## 6.2 Democratic resilience in times of pervasive digital systems

The findings of this dissertation (especially Article V) have revealed a change in the basic structure of democratic societies, which challenges democratic resilience. This challenge is both internal and external: for a nation-state with a democratic regime, providers of pervasive digital systems can originate either inside or outside the country borders.

From the perspective of Rawls's basic structure that should secure social justice, the lack of moral accountability is problematic. If the institutions belonging to the basic structure (e.g., AI system providers) do not align their actions with the principles of justice – as the findings of this dissertation indicate – democratic societies lack in their moral foundations in justice as fairness. As Cohen (2003) notes, in Rawls's theory, justice and democracy are intertangled in a way that one cannot exist without another: fulfilling the principles of justice (e.g., political liberties) is impossible in a non-democratic regime, and collective decision-making (e.g., legislation) is not democratic without fairness. The connection between justice and fairness has been pointed out by several other scholars (e.g., Buchanan, 2002; Pettit, 2012), and the issues around justice and fairness brought forth by AI have been pointed out by, e.g., Coeckelbergh (2024c, pp. 47–48) as one of the ways in which AI undermines democratic principles.

Consequently, the changes in the entity that constitutes the basic structure of society seems to have led to a lack of just background conditions for democratic societies, which is one of the corner stones that for Rawls forms the moral foundations of democratic societies. The current precarity of these foundations can be seen as a step towards these democracies losing their democratic character, which for Merkel and Lührmann (2021) indicates erosion of democratic resilience.

To react and strengthen democratic resilience, institutions belonging to the basic structure would need to collaborate to form an entity that shares the moral duty of securing the just background conditions in democratic societies. Rather than securing their own interests, as can be the tendency of influential technology providers (Westerstrand et al., 2024), regulators, providers of pervasive digital systems, as well as other basic structure institutions should seek for a common direction in steering the development of pervasive digital systems in democratic societies. As Rawls notes, the basic structure is not an individual institution but an "important complex

of institutions", (Rawls, 2005, p. 260) that forms a holistic entity of institutions and interrelations thereof (Rawls, 2005, p. 267). Without such collaboration, fair background conditions might be "gradually undermined even though no one acts unfairly when their conduct is judged by the rules that apply to transactions within the appropriately circumscribed local situations" (Rawls, 2005, p. 267). The risk of such gradual erosion has been noted also in more recent of political theorists, such as Adam Przeworski, who describes the risk as follows:

> "So in the end there are some cases in which the collapse of democracy is manifest, marked by some discrete event, but there are some in which democracy slides down a continuous slope, so not only do we not have discrete markers but we can reasonably disagree about whether a particular regime is still democratic or already past the point of no return." (Przeworski, 2019, p. 26.)

I argue that we are not yet past the point of no return, as the findings of this dissertation indicate willingness from several stakeholders to advance moral duties to develop ethical ISs, as well as to establish organisational governance mechanisms to realise ethical guidance in practice. We have, however, reason to worry, as we could be close to the threshold beyond which we lose our way back to democratic resilience.

For IS researchers, this underlines the need recognised in Chapter 2.1 to move towards an ecosystem perspective that looks at the impacts of technologies from a holistic perspective, recognising their interconnections and the resulting impacts on people, societies and the environment (see, e.g., Stahl, 2021a, 2022). I argue that the findings of this dissertation point towards a need to move from technology-centred ethics perspectives towards a more holistic perspective ethics of digital ecosystems. This can build democratic resilience in a digital age and avoid gradual erosion of democracy by helping different actors to recognise their role in the broader network of institutions that maintain the fundamental concepts of democratic societies, such as justice. I hope that this dissertation and the model it provides (see Figure 10) helps organisations in this endeavour and promote the ability of digital democracies to react to technological changes in a way that supports democratic resilience – if not by resisting change, then by adapting through internal changes, and by recovering without losing the democratic character of their regimes and the constitutive core institutions, organizations, and processes (Merkel & Lührmann, 2021).

As the focus of this dissertation has been biased forwards basic structure institutions and their moral duties, there has not been much discussion about the role of citizens and their agency in this equation. Citizens have only been shortly acknowledged in Article V and the synthesis of the results (Chapter 5.6.4), where it has been stated that there is a need for mechanisms that involve citizens in the

governance efforts in different stages. This choice has been nothing but a practical delimitation that has allowed for this dissertation to retain a reasonable scope. Nevertheless, citizens and their agency in steering the direction of the development of pervasive systems is of high importance and should be kept as an underlying motivation for establishing governance mechanisms that aim towards ethical IS development that supports democratic resilience. Considering that democracy is, by definition, a regime ruled by the people, and that Rawlsian democracy in particular highlights the need for active political agency and deliberation between elections, this is a topic that requires further elaboration in future research.

## 6.3 Contribution to IS research and practice

Apart from the empirical contributions discussed above, part of the research design of this dissertation arises from theoretical motivations. I consider this dissertation to be an initial step towards better understanding and justified critique around pervasive digital systems. With the CPDS methodology and the analytical AIDEM framework, as well as an initial understanding of the AI ethics discourse and its implications to democracy, we can keep accumulating the knowledge and bringing it into practice through AI design and AI governance. This research contributes to the foundational understanding for designing ethical ISs. It can be used to inform the choice of what Young et al. (2024) call a *kernel theory* in their Ethical Design through Grounding and Evaluation (EDGE) – i.e., a theory that is used as a basis for ethical design (Young et al., 2024) – or other artefact-oriented methodologies, such as Design Science Research (Hevner et al., 2004; March & Smith, 1995; Venable, 2010). It can also be used as a stand-alone research approach for accumulating knowledge about ethical and societal implications of pervasive digital systems, as not all the knowledge can and neither should be presented in a form of an artefact.

In the field of IS research, contributing to the understanding of sociotechnical dimensions of ISs is essential, particularly in times of significant technological change. I share the concern expressed by Sarker et al. (2019) on the lack of appreciation towards sociotechnical dimension of IS research. As Sarker et al. (2019) discuss, to ensure the longevity of the field, we need to cultivate the sociotechnical roots of the field. They also distinguish three consequences that losing the ground in sociotechnical perspective can provoke in the field of IS (p. 696):

1. “disciplinary erosion due to a lack of uniqueness;
2. disciplinary fragmentation due to the discipline’s inability to expand in a coherent fashion, and the resulting absence of a shared understanding of topics among its different subcommunities; and

3. a lack of ethical standing of the discipline in society due to the failure of IS scholars and practitioners to reflect on the consequences of information technology, and to critique and actively oppose initiatives where IT might facilitate the development of a dehumanized and dystopian society."

They describe the sociotechnical perspective to be the very *axis of cohesion* for IS research that has shaped the field throughout its history but now seems to be overly neglected by IS researchers (Sarker et al., 2019, p. 670).

Meanwhile, it needs to be recognised that the introduction of ever more advanced algorithmic systems, as well as generative AI interfaces to the market can be seen to have revived the interest towards ethical and societal consequences of ISs. For example, Grover and Lyytinen (2023) call for novel, innovative theories that can address the particularities of pervasive digital systems that are deeply sociotechnical. I argue, however, that rigorous sociotechnical IS research has suffered from the hype around AI systems, with some of the research being conducted due to the pressure of quantified academics (see, Koskinen et al., 2024) to acquire funding and publish papers with contemporary relevance rather than aspiration to cultivate the sociotechnical roots of the IS field. This has resulted in AI ethics research that lacks in rigour when it comes to the ethical justifications of their analyses and artefacts – i.e., the very ethics that the authors have claimed to be in focus. When applied in IS practice, such research can lead to erosion rather than strengthening of ethical awareness in IS research and development (Bleher & Braun, 2023).

This dissertation addresses especially the third point listed by Sarker et al. (2019, p. 696) on the lack of ethical standing by demonstrating ethical reflection in empirical IS contexts. It provides a research approach that facilitates research on the consequences of IS development and enables justified critique towards ISs with negative ethical and/or societal implications. It strengthens the ethical rigour of ethics principles in the context of AI systems and provides a perspective that recognises the role of ISs in societal change in a way that strengthens the continuity of the IS tradition in sociotechnical IS research. I hope this dissertation inspires others, as well, to reject the doubts directed towards moral philosophy and political theory as references for theory development in the field of IS research, and to engage in rigorous sociotechnical IS research by thoroughly reflecting the ethical and societal implications of the systems we develop. Doing so we cultivate the axis of cohesion in IS research and aim towards IS development with positive impact.

## 6.4 General limitations of this dissertation

Besides the above-argued merits, this dissertation comes with several limitations that need to be acknowledged. Whereas some challenges and limitations have been brought forth along with the discussion of the research findings, some remarks remain on the general research design and interpretation of results. One can quickly see that the present dissertation has not offered *complete* answers to any of the three research questions. It has, however, increased our understanding in all fronts and built foundations for further studies that pursue similar goals. The broad scope of the study has come with limitations in depth, which open questions for such further research. What does the AI ethics discourse look like from perspectives of other ethics theories, such as virtue ethics, utilitarian consequentialism, ethics of care, or other deontological theories? How about impacts of specific technologies (e.g., ChatGPT the ethics of which has been analysed by, e.g., Stahl & Eke, 2024), or social media platforms (see Muldoon & Raekstad, 2023)? What kind of discourse is being constructed in media, either journalistic or social media? How different is it in other regulatory contexts, such as the US or China, for instance? These and several other questions remain unanswered, despite being relevant in the construction of AI ethics discourse.

I have only analysed one national AI strategy, one regulation addressing one type of pervasive digital, namely, AI, and taken only one aspect of Rawls's theory under inspection in Article V to study the impacts of AI on democratic societies. In Article I, I have only chosen to look at deliberative democracy, and the analysis therein is arguably succinct. Hence, the contribution to understanding the impacts of pervasive digital systems on democracy remains limited. Moreover, although choosing Rawls as a starting point can be justified (see 4.5), fruitful discussions arise when we engage in comparative analyses. This, however, exceeds the scope of one PhD dissertation, and thus requires further research.

This dissertation also does not take a position regarding one aspect tightly linked to ethical and societal implications of pervasive digital systems: their environmental impacts. Whereas it can be seen as intuitive to think that going digital leads to lesser use of resources, as no physical product is needed to complete a certain task, the rise of pervasive digital systems has raised questions on their strain on the surrounding environment (Bender et al., 2021; Markelius et al., 2024; Rillig et al., 2023). Firstly, algorithms that aim to optimise the use of resources rather lead to an even more efficient use of all available resources than us using less resources. When an organisation optimises one system of extraction, they use the resources freed from that process to extract even more raw materials – something that AI systems have already offered to oil and gas industry (Brevini, 2022, pp. 88–89). Secondly, training and running generative AI models requires significant amounts of computing power (Bender et al., 2021; Brevini, 2022, pp. 74–76; Kanungo, 2023; Ren & Wierman,

2024; Rillig et al., 2023), which has led to companies investing in yet more data centres (Graham, 2024) and even reviving old powerplants (Crownhart, 2024). With the ongoing hype around AI systems this comes with planetary consequences that go hand in hand with the social costs (Markelius et al., 2024). While some environmental benefits can be gained from using AI in individual tasks such as writing and illustrating (Tomlinson et al., 2024), media has reported a tendency for major technology companies to invest in more data centres – sometimes in precarious environments[7] – and even reactivating nuclear plants once shut down due to security incidents[8]. Thus, environmental sustainability is linked to both ethical and societal dimensions.

Moreover, one branch of environmental ethics that takes a non-anthropocentric perspective to ethics argues for intrinsic value of the natural environment (for comparison between the instrumental and intrinsic value perspectives, see Brennan & Lo, 2024). This calls for studying the impacts of pervasive digital on environment for the sake of the intrinsic value of animals other than humans, as well as the environment. In this dissertation, the choice of not addressing this dimension was difficult but pragmatic, arising simply from the need to keep the scope reasonable.

Rawlsian theory offers an interesting avenue for looking at the responsibility of today's societies in regard to future generations, which paves the way for continuing Rawlsian analyses on the impacts of the pervasive digital to an added dimension. As discussed in Chapter 4.5 of this dissertation, Rawls introduces an idea of justice between generations in the form of *just savings principle*. According to this principle, society must agree to a savings principle that "insures that each generation receives its due from its predecessors and does its fair share for those to come" (Rawls, 1999, p. 254). For Rawls, this ensures that "any one generation looks out for all" (Rawls, 1999, p. 255) and requires current generations to ensure institutional stability that allows preservation of the just conditions for people and associations in society. Using this perspective to extend the discussion on intergenerational justice remains a topic for further research.

Moreover, as noted in Chapter 3.1, the constructivist approach assumes the influence of the researcher and their background in the interpretation of research data and findings. In my case, I have received my primary education in a Nordic country

[7] As reported by the Guardian in September 2024, data centre industry is booming in Mexico, where water and electricity is already scarce: https://www.theguardian.com/global-development/2024/sep/25/mexico-datacentre-amazon-google-queretaro-water-electricity. Last accessed: 18 Oct 2024.

[8] According to Reuters (September 2024), Microsoft has signed a deal to resurrect a Three Mile Island nuclear reactor for its increased energy needs: https://www.reuters.com/markets/deals/constellation-inks-power-supply-deal-with-microsoft-2024-09-20/. Last accessed 18 Oct 2024.

and lived almost my entire life in an EU Member state. My prior higher education in Political Science and French language have shaped my thinking even before entering the field of Information Systems research. Therefore, I can only assume that these characteristics have influenced several research choices. I have looked at the research problem from an arguably Eurocentric perspective, which has influenced my choice of the object of research (notably the European AI regulation) and theory, as my Western education has familiarised myself early on with Western political and moral philosophical theory. Although this dissertation does not include a normative position that would indicate an absolute superiority of democracy over other forms of governance – particularly those we are yet to witness – I recognise the bias towards assuming advantages of democratic forms of governance even when some of the theoretical justifications are not present or could be contradictory. Therefore, although these choices have been conscious and a result of extensive theoretical justification discussed in this dissertation, the influence of this background needs to be recognised as a factor that impacts aspects included in and excluded from the scope this study, as well as the interpretations of the research results.

Finally, this dissertation is limited to inspecting the impacts of pervasive digital systems, leaving systems with impactful but less pervasive effects on people, societies and the environment aside. Yet, ISs require governance also when their nature is less pervasive. Moreover, some of the results of this dissertation and the applications of the theoretical background might be different if studied from a perspective of narrower technology in a narrower use case. It is thus recognised that the findings presented here should not be generalised to all ISs and their impacts without critical inspection their characteristics and whether the present research approach responds to the needs arising from those characteristics. This dissertation is conducted using a research approach which is suitable especially for the study of pervasive digital systems and discourses marked by political dimensions. Therefore, systems that do not fall into these categories should be studied separately.

## 6.5 Research agenda: Towards governance of ethical digital ecosystems

The findings and limitations of this dissertation bring forth further gaps in our current knowledge base on the impacts of pervasive digital systems on people, societies and the environment. First, more research is needed to study the impacts from a variety of ethics perspectives, beyond that of Rawls's theory of justice. Besides the valuable contribution that do so in contexts of specific technologies (e.g., Danaher & Nyholm, 2024; Stahl & Eke, 2024), I call for studies that take a broader discourse perspective that taps into the fundamentals of technological change, ethics, and political theory. This requires drawing from moral philosophy to justify the choices of ethics

principles, frameworks and other normative guidance that is used to steer the development of pervasive digital systems to a certain direction. I welcome studies that adopt similar research approach but use, e.g., virtue ethics, consequentialism, or ethics of care as theoretical lenses to defining what direction in technology development is morally justified. For example, the Flourishing Ethics perspective introduced by Bynum (2006) has been shown to be a relevant addition to the study of ethics of contemporary ISs (Kantar & Bynum, 2022; Stahl et al., 2021). Conducting comparative analyses of the ethics discourse of AI in the context of European AI regulation, but through the lens of Flourishing Ethics, would build on the knowledge accumulated in this dissertation and enrichen our understanding of the elements that construct ethics discourses around pervasive digital systems. Considering that Flourishing Ethics also offers a broader lens that exceeds human flourishing, it seems like a viable option for further studies that extends to topics such as environmental sustainability that this dissertation has not yet addressed in depth.

Second, I have revealed a need for integration of ethics into IS practice in a way that avoids the pitfalls of the current principle-based approach and excess focus on technical operationalisation in AI governance. The findings of this dissertation can be used to inform the design of new, innovative AI governance frameworks that come with processes that put ethics into practice as an ongoing process of ethics-informed reflection, deliberation and decision-making. Doing so we can ensure that ethics remains a guiding basis for IS development. Whereas this dissertation has given some indication of necessary considerations in different phases of IS development (see Article IV) and given a framework for providers of pervasive digital systems to take the first steps (see Article V), the next step is to harness methodologies such as Design Science Research to design the processes in collaboration with industry actors, thus increasing the understanding of the nature of the processes and organisational changes needed to put them into practice.

Third, several of the dimensions studied in this dissertation need to be further extended. For example, I have studied AI ethics discourse more deeply than democratic impacts. Therefore, more research is needed from the perspective of different democratic concepts, such as ideal theories and their central characteristics, to increase our understanding of how pervasive digital systems influence elements such as elections (that are the focal point for minimalist democratic theories), public deliberation (central for deliberative democracy and radical democracy), and more. Moreover, to extend the understanding of the AI ethics discourse, we need studies that investigate other contexts than just the EU AI Act or the Spanish AI strategy. As noted, Rawls's theory alone offers more conceptual tools than fit the scope of this dissertation, leaving, e.g., the environmental impacts and intergenerational justice as topics for further research.

Lastly, this dissertation offers an arguably Eurocentric perspective, as the empirical studies focus on the European context of AI regulation and policy. The discussions that go beyond the EU context are Western at best. This has been guided by the research goal of understanding the impacts of pervasive digital system on democratic resilience, which was and still is in need for better understanding. It is not, however, the only one. I recognise that the ethical and societal impacts of pervasive digital systems are just as significant – if not sometimes more so – in other continents and in other political systems. As discussed by, e.g., Couldry and Meijas (2019; 2024), the current extraction logic of data capitalism influences people and ecosystems globally, with the earliest impacts showing in the countries with precarious labour conditions (Perrigo, 2023). The impacts of systems such as AI also come with a range of ethical questions that have not been discussed here when taken into contexts of war or other conflicts, the negative impacts of which have been reported in, e.g., Gaza (e.g., Abraham, 2024). Therefore, there is an urgent need to diversify the perspective towards impacts and discourses that go beyond the European context and seek understanding of the global phenomena, as well as other local phenomena in various corners of the globe.

To illustrate, Figure 11 outlines research topics around the AIDEM framework that still require our attention.

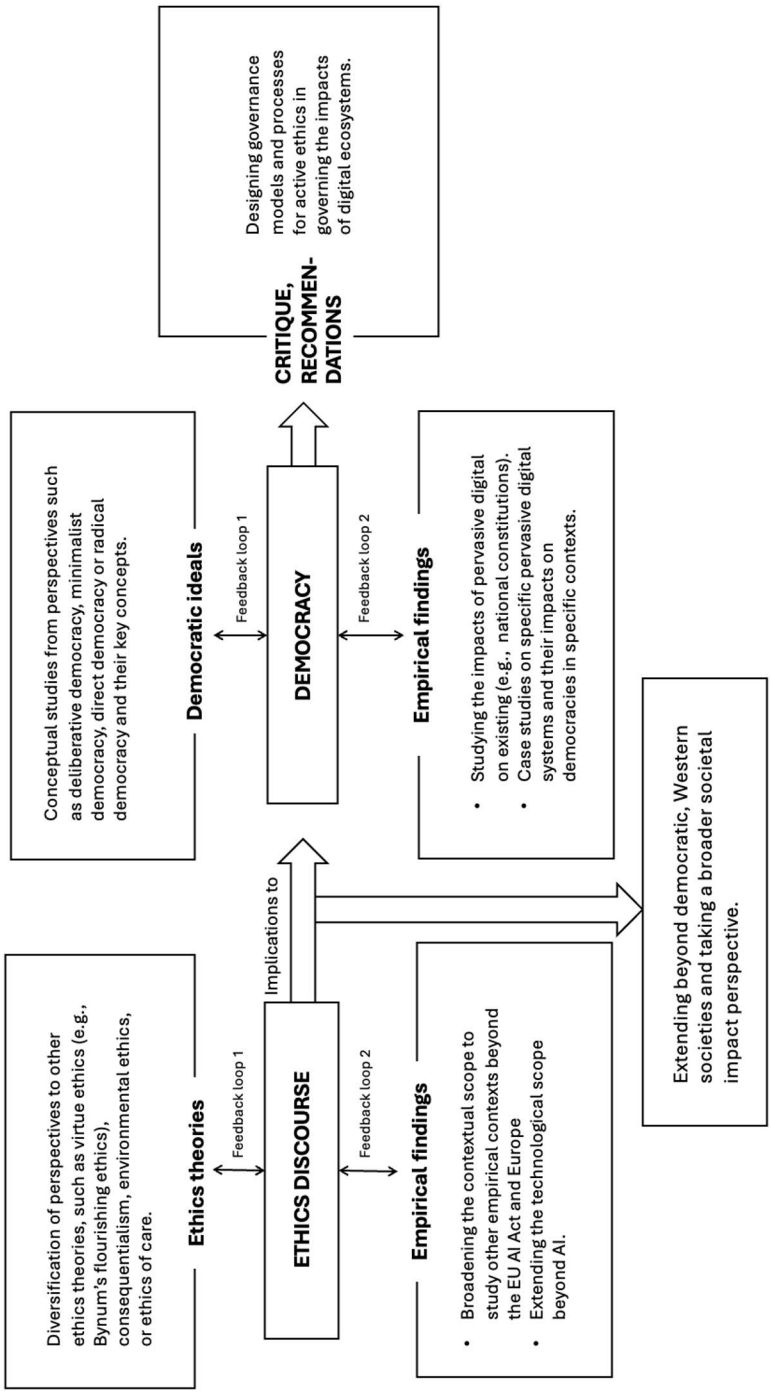


**Figure 11.** Research agenda for ethical governance of digital ecosystems.

# 7 Conclusion

This dissertation studies the ethics discourse around contemporary pervasive digital systems and how it impacts democratic resilience. It is a critical inquiry of the current ways in which we talk about technology, which influence how we develop ISs now and in the future. It is a theoretical and an empirical contribution to broadening the knowledge base about sociotechnical dimensions of ISs and the interplay of these systems with people societies. I have sought answer to three research questions:

- RQ 1: Methodologically, how could we better increase the understanding of sociotechnical dimensions of pervasive digital systems?
- RQ 2: From the perspective of John Rawls's theory of justice, what kinds of ethical dimensions can we distinguish in the ongoing AI ethics discourse?
- RQ 3: In the light of the response to RQ2, what kind of implications do pervasive digital systems have on democratic resilience?

To answer these questions, I have adopted a critical stance and approached the ethics of pervasive digital systems as a discourse. To contribute to IS theory, I have built an analytical framework (Article I) and a Critical-Political Discourse Studies (CPDS) research approach (Article III) that facilitate the study of IS discourses that are marked by a political dimension, which is shown to be a frequent element in contemporary IS phenomena. It relies on critical IS research tradition and the efforts made in studying IS discourses and complements the approach with Political Discourse Analysis methodology. CPDS is thus an innovative research approach that allows IS researchers to better grasp the specificities of IS discourses marked by the political, thus adding to the rigour and depth of the study of contemporary IS phenomena. Together, these frameworks offer a significant contribution to the field of IS theory, enabling access to several of the above-proposed further research questions.

In empirical terms, I have taken Artificial Intelligence (AI) as an example of a type of pervasive digital system and studied the AI ethics discourse produced in the context of AI regulation and national policy in the EU to understand the direction towards which we are heading in AI development and deployment (Article IV). I

have studied the current principle-based approach and contributed to strengthening its basis in moral philosophy (Article II), and analysed how the revealed ethics direction impacts the fundamentals of democratic society (Article V). All this is done through the lens of Rawls's theory of justice as fairness, which serves as a normative theory to define the ethical and societal ideal situation towards which the status quo is compared. To study the AI ethics discourse and to draft a set of guiding principles, I have used Rawls's principles of justice as a backbone and applied them to the context of contemporary AI systems development and deployment. When analysing the democratic implications, I have utilised Rawls's concept of the basic structure of society and steered attention to the impacts the current development and deployment practices have on the ability of the basic structure to secure social justice.

As a result, the main empirical finding of this dissertation follows: from the perspective of John Rawls's theory of justice as fairness, the current way in which we develop and deploy pervasive digital systems – AI in particular – seems to erode rather than support ethical development of pervasive digital systems and democratic resilience. On the level of principles and mission statements, the ongoing AI ethics discourse around European AI regulation takes account of some basic liberties, as well as equality of opportunity. When it comes to the improvement of the situation of the least advantaged members of society, the signs in IS development towards supporting such improvement are weak at best. From a Rawlsian perspective, recognising the moral duty of system providers is required to secure just background conditions for people and associations to function in a democratic society – something that is a fundamental building block of democratic resilience in times of technological change.

We are, however, lacking effective processes to bring ethical reflection into the development processes of ISs, which would be necessary in order for us to guide the development towards ethical direction. Whereas the calls for operationalisation of ethics in practice are not a new finding, this dissertation sheds light on the reasons for which the current approaches have been ineffective, namely, the lack of ethics in AI ethics that serves as a foundation for operationalisation, and the excessive focus on translating principles into technical processes that are eventually distanced from their ethical justifications. Increasing understanding of these phenomena paves the way for meaningful action towards ethical IS development.

As a response, this dissertation proposes a framework for institutions belonging to the basic structure to take steps towards ethical governance of digital ecosystems. It gives guidance in forms of AI ethics principles and a framework laying out considerations for AI system providers throughout systems development to inform the design and business models resulting in pervasive digital systems. It offers a framework for institutions belonging to the basic structure of society to start their journeys towards more ethical and societally sustainable digital development. To

bring these artefacts together, I propose a model for a resilient basic structure of society that guarantees just background conditions for individuals and associations to function in the context of pervasive digital systems (see Figure 10 above).

In addition, this dissertation has revealed several questions for further research (see Figure 11 above) that should be addressed to continue the efforts in building understanding and offering guidance for action in the industry. Among these, we should strengthen the basis of IS ethics and governance studies in moral philosophy, extend the perspective from Rawls to other theorists, and go beyond the European regulation to study the variety of discourses around the globe. Doing so we could gain a fuller picture of the direction to which we are currently heading and offer critique and recommendations based on rigorous theory basis and insightful empirical studies. This dissertation implies that doing so would require broadening the perspective from individual technologies also to studying the interconnections between technologies and their joint impacts on broader ecosystems. I thus invite researchers in the field of IS and beyond to join me in efforts to continue contributing to the critical study of impacts of pervasive digital systems on people, societies and the environment – in Europe and beyond.

# Abbreviations

| | |
|---|---|
| AI | Artificial Intelligence |
| CDS | Critical Discourse Studies |
| CPDS | Critical-Political Discourse Studies |
| EU AI Act | Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) |
| GenAI | Generative Artificial Intelligence |
| GPAI | General-Purpose Artificial Intelligence |
| IS | Information Systems |
| PDA | Political Discourse Analysis |
| RQ | Research question |

# List of References


Abraham, Y. (2024, April 3). *'Lavender': The AI machine directing Israel's bombing spree in Gaza*. https://www.972mag.com/lavender-ai-israeli-army-gaza/. Last accessed 25 Jan 2025.

Adams, R. (2021). Can artificial intelligence be decolonized? *Interdisciplinary Science Reviews*, *46*(1–2), 176–197. https://doi.org/10.1080/03080188.2020.1840225.

Adorno, T. W., & Horkheimer, M. (1979). *Dialectic of Enlightenment*. Verso.

Albertus, R. W., & Makoza, F. (2023). Habermasian analysis of reports on Presidential tweets influencing politics in the USA. *International Politics*, *60*(2), 330–349. https://doi.org/10.1057/s41311-022-00396-7.

Alnemr, N. (2020). Emancipation cannot be programmed: Blind spots of algorithmic facilitation in online deliberation. *Contemporary Politics*, *26*(5), 531–552. https://doi.org/10.1080/13569775.2020.1791306.

Alvarez, R. (2002). Confessions of an information worker: A critical analysis of information requirements discourse. *Information and Organization*, *12*(2), 85–107. https://doi.org/10.1016/S1471-7727(01)00012-4.

Alvarez, R. (2005). Talking a critical linguistic turn: Using critical discourse analysis for the study of information systems. In D. Howcroft & E. M. Trauth (Eds.), *Handbook of Critical Information Systems Research: Theory and Application* (pp. 104–122). Edward Elgar Publishing.

Alvesson, M., & Deetz, S. (2000). *Doing Critical Management Research*. SAGE.

Alvesson, M., & Willmott, H. (1992). On the Idea of Emancipation in Management and Organization Studies. *Academy of Management Review*, *17*(3), 432–464. https://doi.org/10.5465/amr.1992.4281977.

Angermüller, J. (2011). Heterogeneous knowledge: Trends in German discourse analysis against an international background. *Journal of Multicultural Discourses*, *6*(2), 121–136. https://doi.org/10.1080/17447143.2011.582117.

Auramäki, E., Hirschheim, R., & Lyytinen, K. (1992). Modelling Offices Through Discourse Analysis: The SAMPO Approach. *The Computer Journal*, *35*(4), 342–352. https://doi.org/10.1093/comjnl/35.4.342.

Aylsworth, T., & Castro, C. (2024). *Kantian Ethics and the Attention Economy*. https://doi.org/10.1007/978-3-031-45638-1.

Bates, D. W., Levine, D., Syrowatka, A., Kuznetsova, M., Craig, K. J. T., Rui, A., Jackson, G. P., & Rhee, K. (2021). The potential of artificial intelligence to improve patient safety: A scoping review. *Npj Digital Medicine*, *4*(1), 1–8. https://doi.org/10.1038/s41746-021-00423-6.

Bednar, K., & Spiekermann, S. (2024). The Power of Ethics: Uncovering Technology Risks and Positive Value Potentials in IT Innovation Planning. *Business & Information Systems Engineering*, *66*(2), 181–201. https://doi.org/10.1007/s12599-023-00837-4.

Beil, M., Proft, I., van Heerden, D., Sviri, S., & van Heerden, P. V. (2019). Ethical considerations about artificial intelligence for prognostication in intensive care. *Intensive Care Medicine Experimental*, *7*(1), 70. https://doi.org/10.1186/s40635-019-0286-6.

Benbya, H., Nan, N., Tanriverdi, H., & Yoo, Y. (2020). Complexity and Information Systems Research in the Emerging Digital World. *MIS Quarterly,* 44(1), 1–17. DOI: 10.25300/MISQ/2020/13304.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922.

Bennett, M. R., & Hacker, P. M. S. (2021). *Philosophical Foundations of Neuroscience*. John Wiley & Sons.

Berkey, B. (2021). Rawlsian Institutionalism and Business Ethics: Does It Matter Whether Corporations Are Part of the Basic Structure of Society? *Business Ethics Quarterly*, *31*(2), 179–209. https://doi.org/10.1017/beq.2020.14.

Bermeo, N. (2016). On Democratic Backsliding. *Journal of Democracy*, *27*(1), 5–19.

Berry, D. M. (2014). *Critical Theory and the Digital*. A&C Black.

Bietti, E. (2021). From Ethics Washing to Ethics Bashing: A Moral Philosophy View on Tech Ethics. *Journal of Social Computing*, *2*(3), 266–283. https://doi.org/10.23919/JSC.2021.0031.

Birkstedt, T., Minkkinen, M., Tandon, A., & Mäntymäki, M. (2023). AI governance: Themes, knowledge gaps and future agendas. *Internet Research*, *33*(7), 133–167. https://doi.org/10.1108/INTR-01-2022-0042.

Blanc, S., & Al-Amoudi, I. (2013). Corporate Institutions in a Weakened Welfare State: A Rawlsian Perspective. *Business Ethics Quarterly*, *23*(4), 497–525. https://doi.org/10.5840/beq201323438.

Bleher, H., & Braun, M. (2023). Reflections on Putting AI Ethics into Practice: How Three AI Ethics Approaches Conceptualize Theory and Practice. *Science and Engineering Ethics*, *29*(3), 21. https://doi.org/10.1007/s11948-023-00443-3.

Bostrom, N. (2017). *Superintelligence: Paths, Dangers, Strategies*. Dunod.

Bouvier, G., & Machin, D. (2020). Critical Discourse Analysis and the challenges and opportunities of social media. In *Critical Discourse Studies and/in Communication*. Routledge.

Braun, M., & Meacham, D. (2024). A Plea for (In)Human-centred AI. *Philosophy & Technology*, *37*(3), 97. https://doi.org/10.1007/s13347-024-00785-1.

Brennan, A., & Lo, N. Y. S. (2024). Environmental Ethics. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Summer 2024). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2024/entries/ethics-environmental/.

Brevini, B. (2022). *Is AI Good for the Planet?* Polity Press.

Brkan, M. (2019). Artificial Intelligence and Democracy: *Delphi - Interdisciplinary Review of Emerging Technologies*, *2*(2), 66–71. https://doi.org/10.21552/delphi/2019/2/4.

Brock, A. (2018). Critical technocultural discourse analysis. *New Media & Society*, *20*(3), 1012–1030. https://doi.org/10.1177/1461444816677532.

Buchanan, A. (2002). Political Legitimacy and Democracy. *Ethics*, *112*(4), 689–719. https://doi.org/10.1086/340313.

Bynum, T. W. (2006). Flourishing Ethics. *Ethics and Information Technology*, *8*(4), 157–173. https://doi.org/10.1007/s10676-006-9107-1.

Bynum, T. W. (2008). Norbert Wiener and the Rise of Information Ethics. In M. J. van den Joven & J. Weckert (Eds.), *Information Technology and Moral Philosophy*. Cambridge University Press.

Card, D., & Smith, N. A. (2020). On Consequentialism and Fairness. *Frontiers in Artificial Intelligence*, *3*. https://doi.org/10.3389/frai.2020.00034.

Cecez-Kecmanovic, D., Davison, R., Fernandez, W., Finnegan, P., Pan, S., & Sarker, S. (2020). Advancing Qualitative IS Research Methodologies: Expanding Horizons and Seeking New Paths. *Journal of the Association for Information Systems*, *21*(1). https://doi.org/10.17705/1jais.00599.

Chandler, D. (2023). *Free and Equal: What Would a Fair Society Look Like?* Penguin UK.

Charmaz, K. (2020). "With Constructivist Grounded Theory You Can't Hide": Social Justice Research and Critical Inquiry in the Public Sphere. *Qualitative Inquiry*, *26*(2), 165–176. https://doi.org/10.1177/1077800419879081.

Chesney, B., & Citron, D. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*, *107*(6), 1753–1820.

Chiasson, M., Davidson, E., & Winter, J. (2018). Philosophical foundations for informing the future(S) through IS research. *European Journal of Information Systems*, *27*(3), 367–379. https://doi.org/10.1080/0960085X.2018.1435232.

Clarke, R. (2020). The Challenges Involved in Establishing a Research Technique. *Australasian Journal of Information Systems*, *24*. https://doi.org/10.3127/ajis.v24i0.2515.

Cocchiaro, M. Z., Morley, J., Novelli, C., Panai, E., Tartaro, A., & Floridi, L. (2024). *Who is an AI Ethicist? An Empirical Study of Expertise, Skills, and Profiles to Build a Competency Framework* (SSRN Scholarly Paper 4891907). https://doi.org/10.2139/ssrn.4891907

Coeckelbergh, M. (2022). *The Political Philosophy of AI: An Introduction*. John Wiley & Sons.

Coeckelbergh, M. (2024a). All too real metacapitalism: Towards a non-dualist political ontology of metaverse. *Ethics and Information Technology*, *26*(2), 30. https://doi.org/10.1007/s10676-024-09768-4.

Coeckelbergh, M. (2024b). Artificial intelligence, the common good, and the democratic deficit in AI governance. *AI and Ethics*. https://doi.org/10.1007/s43681-024-00492-9.

Coeckelbergh, M. (2024c). *Why AI Undermines Democracy and What To Do About It*. John Wiley & Sons.

Cohen, J. (2003). For a Democratic Society. In S. Freeman (Ed.), *The Cambridge Companion to Rawls* (pp. 86–167). Cambridge University Press.

Congge, U., Guillamón, M.-D., Nurmandi, A., Salahudin, & Sihidi, I. T. (2023). Digital democracy: A systematic literature review. *Frontiers in Political Science*, *5*. https://doi.org/10.3389/fpos.2023.972802.

Constantinescu, M., & Crisp, R. (2022). Can Robotic AI Systems Be Virtuous and Why Does This Matter? *International Journal of Social Robotics*, *14*(6), 1547–1557. https://doi.org/10.1007/s12369-022-00887-w.

Couldry, N., & Mejias, U. A. (2019). Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. *Television & New Media*, *20*(4), 336–349. https://doi.org/10.1177/1527476418796632.

Crownhart, C. (2024, September 26). Why Microsoft made a deal to help restart Three Mile Island. *MIT Technology Review*. https://www.technologyreview.com/2024/09/26/1104516/three-mile-island-microsoft/.

Cukier, W., Ngwenyama, O., Bauer, R., & Middleton, C. (2009). A critical analysis of media discourse on information technology: Preliminary results of a proposed method for critical discourse analysis. *Information Systems Journal*, *19*(2), 175–196. https://doi.org/10.1111/j.1365-2575.2008.00296.x.

Danaher, J., & Nyholm, S. (2024). Digital Duplicates and the Scarcity Problem: Might AI Make Us Less Scarce and Therefore Less Valuable? *Philosophy & Technology*, *37*(3), 106. https://doi.org/10.1007/s13347-024-00795-z.

De Moya, J.-F., & Pallud, J. (2020). From panopticon to heautopticon: A new form of surveillance introduced by quantified-self practices. *Information Systems Journal*, *30*(6), 940–976. https://doi.org/10.1111/isj.12284.

Deetz, S. (1996). Crossroads—Describing Differences in Approaches to Organization Science: Rethinking Burrell and Morgan and Their Legacy. *Organization Science*, *7*(2), 191–207. https://doi.org/10.1287/orsc.7.2.191.

Delanty, G., & Harris, N. (2021). Critical theory and the question of technology: The Frankfurt School revisited. *Thesis Eleven*, *166*(1), 88–108. https://doi.org/10.1177/07255136211002055.

Dhar, P. (2020). The carbon impact of artificial intelligence. *Nature Machine Intelligence*, *2*(8), 423–425. https://doi.org/10.1038/s42256-020-0219-9.

Diakopoulos, N., & Johnson, D. (2021). Anticipating and addressing the ethical implications of deepfakes in the context of elections. *New Media & Society*, *23*(7), 2072–2098. https://doi.org/10.1177/1461444820925811.

Dolata, M., & Schwabe, G. (2023). What is the Metaverse and who seeks to define it? Mapping the site of social construction. *Journal of Information Technology*, *38*(3), 239–266. https://doi.org/10.1177/02683962231159927.

Douglas, D. M. (2015). Towards a just and fair Internet: Applying Rawls' principles of justice to Internet regulation. *Ethics and Information Technology*, *17*(1), 57–64. https://doi.org/10.1007/s10676-015-9361-1.

Du, H., Teng, S., Chen, H., Ma, J., Wang, X., Gou, C., Li, B., Ma, S., Miao, Q., Na, X., Ye, P., Zhang, H., Luo, G., & Wang, F.-Y. (2023). Chat With ChatGPT on Intelligent Vehicles: An IEEE TIV Perspective. *IEEE Transactions on Intelligent Vehicles*, *8*(3), 2020–2026. IEEE Transactions on Intelligent Vehicles. https://doi.org/10.1109/TIV.2023.3253281.

Fairclough, I., & Fairclough, N. (2013). *Political Discourse Analysis: A Method for Advanced Students*. Routledge.

Farina, M., Zhdanov, P., Karimov, A., & Lavazza, A. (2024). AI and society: A virtue ethics approach. *AI & SOCIETY*, *39*(3), 1127–1140. https://doi.org/10.1007/s00146-022-01545-5.

Farrell, H. (2012). The Consequences of the Internet for Politics. *Annual Review of Political Science*, *15*(Volume 15, 2012), 35–52. https://doi.org/10.1146/annurev-polisci-030810-110815

Feenberg, A. (1991). *Critical Theory of Technology* (J. K. B. O. Friis, S. A. Pedersen, & V. F. Hendricks, Eds.). Oxford University Press.

Feenberg, A. (1999). *Questioning Technology*. Taylor & Francis Group.

Fleuß, D., & Schaal, G. S. (2019). *What Are We Doing When We Are Doing Democratic Theory?* https://doi.org/10.3167/dt.2019.060203.

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.

Formosa, P. (2021). Robot Autonomy vs. Human Autonomy: Social Robots, Artificial Intelligence (AI), and the Nature of Autonomy. *Minds and Machines*, *31*(4), 595–616. https://doi.org/10.1007/s11023-021-09579-2.

Franke, U. (2021). Rawls's Original Position and Algorithmic Fairness. *Philosophy & Technology*, *34*(4), 1803–1817. https://doi.org/10.1007/s13347-021-00488-x.

Franke, U. (2024). Rawlsian Algorithmic Fairness and a Missing Aggregation Property of the Difference Principle. *Philosophy & Technology*, *37*(3), 87. https://doi.org/10.1007/s13347-024-00779-z.

Franzke, A. S. (2022). An exploratory qualitative analysis of AI ethics guidelines. *Journal of Information, Communication and Ethics in Society*, *20*(4), 401–423. https://doi.org/10.1108/JICES-12-2020-0125.

Gabriel, I. (2022). Toward a Theory of Justice for Artificial Intelligence. *Daedalus*, *151*(2), 218–231.

Goede, R., & Boshuizen-van Burken, C. (2019). A critical systems thinking approach to empower refugees based on Maslow's theory of human motivation. *Systems Research and Behavioral Science*, *36*(5), 715–726. https://doi.org/10.1002/sres.2623.

Goede, R., & Taylor, E. (2019). Theory in Emancipative Action: Aligning Action Research in Information Systems Education with Critical Social Research in Information Systems. *Systems*, *7*(3), Article 3. https://doi.org/10.3390/systems7030036.

Graham, T. (2024, September 25). Mexico's datacentre idustry is booming—But are more drought and blackouts the price communities must pay? *The Guardian*. https://www.theguardian.com/global-development/2024/sep/25/mexico-datacentre-amazon-google-queretaro-water-electricity. Last accessed 25 Jan 2025.

Grillo, E., Luo, Z., Nalepa, M., & Prato, C. (2024). Theories of Democratic Backsliding. *Annual Review of Political Science*, *27*, 381–400. https://doi.org/10.1146/annurev-polisci-041322-025352.

Grote, T. (2022). Randomised controlled trials in medical AI: Ethical considerations. *Journal of Medical Ethics*, *48*(11), 899–906. https://doi.org/10.1136/medethics-2020-107166.

Grover, V., & Lyytinen, K. (2023). The Pursuit of Innovative Theory in the Digital Age. *Journal of Information Technology*, *38*(1), 45–59. https://doi.org/10.1177/02683962221077112.

Grover, V., & Niederman, F. (2021). Research Perspectives: The Quest for Innovation in Information Systems Research: Recognizing, Stimulating, and Promoting Novel and Useful Knowledge. *Journal of the Association for Information Systems*, *22*(6), 1753–1782. https://doi.org/10.17705/1jais.00705.

Habermas, J. (1976). *Theory and practice*. Polity Press.

Habermas, J. (1984). *The theory of communicative action, Volume 1* (Vol. 1). Polity Press.

Habermas, J. (1996). *Between facts and norms. Transl. William Rehg*. Polity.

Hacker, K. L., & Dijk, J. van. (2000). *Digital Democracy: Issues of Theory and Practice*. SAGE.

Hacker, P. (2018). Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law. *Common Market Law Review*, *55*(4), 1143–1185. https://doi.org/10.54648/cola2018095.

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, *30*(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8.

Hansen, P., & Caterino, B. (2019). *Critical Theory, Democracy, and the Challenge of Neo-Liberalism*. University of Toronto Press.

Harsanyi, J. C. (1975). Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory. *American Political Science Review*, *69*(2), 594–606. https://doi.org/10.2307/1959090.

Heidari, H., Loi, M., Gummadi, K. P., & Krause, A. (2019). A Moral Framework for Understanding Fair ML through Economic Models of Equality of Opportunity. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 181–190. https://doi.org/10.1145/3287560.3287584.

Heilinger, J.-C. (2022). The Ethics of AI Ethics. A Constructive Critique. *Philosophy & Technology*, *35*(3), 61. https://doi.org/10.1007/s13347-022-00557-9.

Heimo, O. I., Fairweather, N. B., & Kimppa, K. K. (2010). The Finnish e-voting experiment: What went wrong. *The 'Backwards, Forwards and Sideways' Changes of ICT. ETHICOMP Proceedings 2010*, 290.

Helbing, D. (2021). *Next Civilization: Digital Democracy and Socio-Ecological Finance-How to Avoid Dystopia and Upgrade Society by Digital Means*. Springer Nature.

Hermann, E. (2022). Leveraging Artificial Intelligence in Marketing for Social Good—An Ethical Perspective. *Journal of Business Ethics*, *179*(1), 43–61. https://doi.org/10.1007/s10551-021-04843-y.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75–105. https://doi.org/10.2307/25148625.

Hirschheim, R., & Klein, H. K. (1994). Realizing Emancipatory Principles in Information Systems Development: The Case for ETHICS. *MIS Quarterly*, *18*(1), 83–109. https://doi.org/10.2307/249611.

Hirschheim, R., Klein, H. K., & Lyytinen, K. (1995). *Information Systems Development and Data Modeling: Conceptual and Philosophical Foundations*. Cambridge University Press.

Hollanek, T. (2023). AI transparency: A matter of reconciling design with critique. *AI & SOCIETY*, *38*(5), 2071–2079. https://doi.org/10.1007/s00146-020-01110-y.

Holloway, J., & Manwaring, R. (2023). How well does 'resilience' apply to democracy? A systematic review. *Contemporary Politics*, *29*(1), 68–92. https://doi.org/10.1080/13569775.2022.2069312.

Hunkenschroer, A. L., & Kriebitz, A. (2023). Is AI recruiting (un)ethical? A human rights perspective on the use of AI for hiring. *AI and Ethics*, *3*(1), 199–213. https://doi.org/10.1007/s43681-022-00166-4.

Hunter, L. (2024). Compulsion beyond fairness: Towards a critical theory of technological abstraction in neural networks. *AI & SOCIETY*. https://doi.org/10.1007/s00146-024-02035-6.

Hur, I., Cousins, K. C., & Stahl, B. C. (2019). A critical perspective of engagement in online health communities. *European Journal of Information Systems*, *28*(5), 523–548. https://doi.org/10.1080/0960085X.2019.1620477.

Ibáñez, J. C., & Olmeda, M. V. (2022). Operationalising AI ethics: How are companies bridging the gap between practice and principles? An exploratory study. *AI & SOCIETY*, *37*(4), 1663–1687. https://doi.org/10.1007/s00146-021-01267-0.

Jacobs, N., & Huldtgren, A. (2021). Why value sensitive design needs ethical commitments. *Ethics and Information Technology*, *23*(1), 23–26. https://doi.org/10.1007/s10676-018-9467-3.

Javaid, M., Haleem, A., & Singh, R. P. (2023). ChatGPT for healthcare services: An emerging stage for an innovative perspective. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, *3*(1), 100105. https://doi.org/10.1016/j.tbench.2023.100105.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), Article 9. https://doi.org/10.1038/s42256-019-0088-2.

Jones, M. (1997). It All Depends What You Mean by Discipline... In J. Mingers & F. A. Stowell (Eds.), *Information Systems: An Emerging Discipline?* (pp. 97–110). McGraw-Hill.

Jungherr, A. (2023). Artificial Intelligence and Democracy: A Conceptual Framework. *Social Media + Society*, *9*(3), 20563051231186353. https://doi.org/10.1177/20563051231186353.

Jungherr, A., & Schroeder, R. (2023). Artificial intelligence and the public arena. *Communication Theory*, *33*(2–3), 164–173. https://doi.org/10.1093/ct/qtad006.

Kane, K., Young, A., Majchrzak, A., & Ransbotham, S. (2021). Avoiding an Oppressive Future of Machine Learning: A Design Theory for Emancipatory Assistants. *Management Information Systems Quarterly*, *45*(1), 371–396.

Kantar, N., & Bynum, T. W. (2022). Flourishing Ethics and identifying ethical values to instill into artificially intelligent agents. *Metaphilosophy*, *53*(5), 599–604. https://doi.org/10.1111/meta.12583.

Kanungo, A. (2023, July 18). *The Green Dilemma: Can AI Fulfil Its Potential Without Harming the Environment?* Earth.Org. https://earth.org/the-green-dilemma-can-ai-fulfil-its-potential-without-harming-the-environment/. Last accessed 25 Jan 2025.

Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Patterns*, *2*(9). https://doi.org/10.1016/j.patter.2021.100314.

Keeling, G. (2018). Against Leben's Rawlsian Collision Algorithm for Autonomous Vehicles. In V. C. Müller (Ed.), *Philosophy and Theory of Artificial Intelligence 2017* (pp. 259–272). Springer International Publishing. https://doi.org/10.1007/978-3-319-96448-5_29.

Kelly, M. (1994). *Critique and Power: Recasting the Foucault/Habermas Debate*. MIT Press.

Kilovaty, I. (2019). Legally cognizable manipulation. *Berkeley Tech. LJ*, *34*, 449.

Koniakou, V. (2023). From the "rush to ethics" to the "race for governance" in Artificial Intelligence. *Information Systems Frontiers*, *25*(1), 71–102. https://doi.org/10.1007/s10796-022-10300-6.

Koskinen, J., Kimppa, K. K., Lahtiranta, J., & Hyrynsalmi, S. (2024). Quantified academics: Heideggerian technology critical analysis of the academic ranking competition. *Information Technology & People*, *37*(8), 25–42. https://doi.org/10.1108/ITP-01-2023-0032.

Kreps, S., & Kriner, D. (2023). How AI Threatens Democracy. *Journal of Democracy*, *34*(4), 122–131.

Krijger, J. (2022). Enter the metrics: Critical theory and organizational operationalization of AI ethics. *AI & SOCIETY*, *37*(4), 1427–1437. https://doi.org/10.1007/s00146-021-01256-3.

Łabuz, M., & Nehring, C. (2024). On the way to deep fake democracy? Deep fakes in election campaigns in 2023. *European Political Science*. https://doi.org/10.1057/s41304-024-00482-9.

Landemore, H. (2021). Open Democracy and Digital Technologies. In L. Bernholz, H. Landemore, & R. Reich (Eds.), *Digital Technology and Democratic Theory* (pp. 62–89). The University of Chicago Press.

Leben, D. (2017). A Rawlsian algorithm for autonomous vehicles. *Ethics and Information Technology*, *19*(2), 107–115. https://doi.org/10.1007/s10676-017-9419-3.

Leben, D. (2018). *Ethics for Robots: How to Design a Moral Algorithm*. Routledge. https://doi.org/10.4324/9781315197128.

Livingston, S., & Risse, M. (2019). The Future Impact of Artificial Intelligence on Humans and Human Rights. *Ethics & International Affairs*, *33*(2), 141–158. https://doi.org/10.1017/S089267941900011X.

Loader, B., & Mercea, D. (2012). *Social Media and Democracy: Innovations in Participatory Politics*. Routledge.

Mainz, J. T., Sønderholm, J., & Uhrenfeldt, R. (2024). Artificial intelligence and the secret ballot. *AI & SOCIETY*, *39*(2), 515–522. https://doi.org/10.1007/s00146-022-01551-7.

Manheim, K., & Kaplan, L. (2019). Artificial Intelligence: Risks to Privacy and Democracy. *Yale Journal of Law & Technology,* 21, 106.

Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022). Defining organizational AI governance. *AI and Ethics*, *2*(4), 603–609. https://doi.org/10.1007/s43681-022-00143-x.

Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2023). *Putting AI Ethics into Practice: The Hourglass Model of Organizational AI Governance* (arXiv:2206.00335). arXiv. https://doi.org/10.48550/arXiv.2206.00335.

Mäntymäki, M., Minkkinen, M., Zimmer, M., Birkstedt, T., & Viljanen, M. (2023). Designing an AI governance framework: From research-based premises to meta-requirements. *ECIS 2023 Research Papers*. https://aisel.aisnet.org/ecis2023_rp/295.

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, *15*(4), 251–266. https://doi.org/10.1016/0167-9236(94)00041-2.

Markelius, A., Wright, C., Kuiper, J., Delille, N., & Kuo, Y.-T. (2024). The mechanisms of AI hype and its planetary and social costs. *AI and Ethics*, *4*(3), 727–742. https://doi.org/10.1007/s43681-024-00461-2.

Masiero, S. (2023). Decolonising critical information systems research: A subaltern approach. *Information Systems Journal*, *33*(2), 299–323. https://doi.org/10.1111/isj.12401.

Mazière, F. (2005). *L'Analyse du discours: Histoire et pratiques*. Presses Universitaires de France.

McKenna, B., & Chughtai, H. (2020). Resistance and sexuality in virtual worlds: An LGBT perspective. *Computers in Human Behavior*, *105*, 106199. https://doi.org/10.1016/j.chb.2019.106199.

McLeod, S. K., & Tanyi, A. (2021). The basic liberties: An essay on analytical specification. *European Journal of Political Theory*, 14748851211041702. https://doi.org/10.1177/14748851211041702.

Meaker, M. (2023, October 3). Slovakia's Election Deepfakes Show AI Is a Danger to Democracy. *Wired*. https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy/. Last accessed 25 Jan 2025.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2022). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, *54*(6), 1–35. https://doi.org/10.1145/3457607.

Mejias, U. A., & Couldry, N. (2024). *Data Grab: The new Colonialism of Big Tech and how to fight back*. Random House.

Merkel, W., & Lührmann, A. (2021). Resilience of democracies: Responses to illiberal and authoritarian challenges. *Democratization*, *28*(5), 869–884.

Miller, A. (2003). *An Introduction to Contemporary Metaethics*. Polity.

Miller, B. (2021). Is Technology Value-Neutral? *Science, Technology, & Human Values*, *46*(1), 53–80. https://doi.org/10.1177/0162243919900965.

Mingers, J., & Standing, C. (2020). A Framework for Validating Information Systems Research Based on a Pluralist Account of Truth and Correctness. *Journal of the Association for Information Systems*, *21*(1). https://doi.org/10.17705/1jais.00594.

Mingers, J., & Walsham, G. (2010). Toward Ethical Information Systems: The Contribution of Discourse Ethics. *MIS Quarterly*, *34*(4), 833–854. https://doi.org/10.2307/25750707.

Mitchell, S., Potash, E., Barocas, S., D'Amour, A., & Lum, K. (2021). Algorithmic Fairness: Choices, Assumptions, and Definitions. *Annual Review of Statistics and Its Application*, *8*(1), 141–163. https://doi.org/10.1146/annurev-statistics-042720-125902.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, *1*(11), Article 11. https://doi.org/10.1038/s42256-019-0114-4.

Monson, M. (2023). Socially responsible design science in information systems for sustainable development: A critical research methodology. *European Journal of Information Systems*, *32*(2), 207–237. https://doi.org/10.1080/0960085X.2021.1946442.

Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mökander, J., & Floridi, L. (2021). Ethics as a Service: A Pragmatic Operationalisation of AI Ethics. *Minds and Machines*, *31*(2), 239–256. https://doi.org/10.1007/s11023-021-09563-w.

Morley, J., Kinsey, L., Elhalal, A., Garcia, F., Ziosi, M., & Floridi, L. (2023). Operationalising AI ethics: Barriers, enablers and next steps. *AI & SOCIETY*, *38*(1), 411–423. https://doi.org/10.1007/s00146-021-01308-8.

Muldoon, J. (2022). *Platform Socialism: How to Reclaim our Digital Future from Big Tech*. Pluto Press.

Muldoon, J., & Raekstad, P. (2023). Algorithmic domination in the gig economy. *European Journal of Political Theory*, *22*(4), 587–607. https://doi.org/10.1177/14748851221082078.

Mumford, E. (1983). *Designing Human Systems for New Technology: The ETHICS Method*. Manchester Business School.

Mumford, E. (1998). Problems, Knowledge, Solutions: Solving Complex Problems. *ICIS Research Papers*. https://dl.acm.org/doi/pdf/10.5555/353053.353134.

Mumford, E. (2003). *Redesigning Human Systems*. Idea Group Inc (IGI).

Myers, M. D., & Klein, H. K. (2011). A Set of Principles for Conducting Critical Research in Information Systems. *MIS Quarterly*, *35*(1), 17–36. https://doi.org/10.2307/23043487.

Nemitz, P. (2018). Constitutional democracy and technology in the age of artificial intelligence. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2133), 20180089. https://doi.org/10.1098/rsta.2018.0089.

Ngwenyama, O., Rowe, F., Klein, S., & Henriksen, H. Z. (2023). The Open Prison of the Big Data Revolution: False Consciousness, Faustian Bargains, and Digital Entrapment. *Information Systems Research*. https://doi.org/10.1287/isre.2020.0588.

Nozick, R. (2013). *Anarchy, State, and Utopia*. Basic Books.

Orlikowski, W. J. (1992). The Duality of Technology: Rethinking the Concept of Technology in Organizations. *Organization Science*, *3*(3), 398–427. https://doi.org/10.1287/orsc.3.3.398.

Orlikowski, W. J., & Baroudi, J. J. (1991). Studying Information Technology in Organizations: Research Approaches and Assumptions. *Information Systems Research*, *2*(1), 1–28. https://doi.org/10.1287/isre.2.1.1.

Paterson, T., & Hanley, L. (2020). Political warfare in the digital age: Cyber subversion, information operations and 'deep fakes'. *Australian Journal of International Affairs*, *74*(4), 439–454. https://doi.org/10.1080/10357718.2020.1734772.

Perrigo, B. (2023, January 18). Exclusive: OpenAI Used Kenyan Workers on Less Thank $2 Per Hour to Make ChatGPT Less Toxic. *Time Magazine*. https://time.com/6247678/openai-chatgpt-kenya-workers/. Last accessed 25 Jan 2025.

Persily, N., Tucker, J. A., & Tucker, J. A. (2020). *Social Media and Democracy: The State of the Field, Prospects for Reform*. Cambridge University Press.

Pettit, P. (2012). *On the People's Terms*. Cambridge University Press.

Pizzi, M., Romanoff, M., & Engelhardt, T. (2020). AI for humanitarian action: Human rights and ethics. *International Review of the Red Cross*, *102*(913), 145–180. https://doi.org/10.1017/S1816383121000011.

Porra, J., & Hirschheim, R. (2007). A Lifetime of Theory and Action on the Ethical Use of Computers: A Dialogue with Enid Mumford. *Journal of the Association for Information Systems*, *8*(9), 3.

Pozzebon, M., & Pinsonneault, A. (2005). Global–local negotiations for implementing configurable packages: The power of initial organizational decisions. *The Journal of Strategic Information Systems*, *14*(2), 121–145. https://doi.org/10.1016/j.jsis.2005.04.004.

Price, E. (2013). Social media and democracy. *Australian Journal of Political Science*, *48*(4), 519–527. https://doi.org/10.1080/10361146.2013.846296.

Prunkl, C. (2024). Human Autonomy at Risk? An Analysis of the Challenges from AI. *Minds and Machines*, *34*(3), 26. https://doi.org/10.1007/s11023-024-09665-1.

Przeworski, A. (2019). *Crises of Democracy*. Cambridge University Press.

Raab, C. D. (2020). Information privacy, impact assessment, and the place of ethics. *Computer Law & Security Review*, *37*, 105404. https://doi.org/10.1016/j.clsr.2020.105404.

Rasmussen, S. L., & Sahay, S. (2021). Engaging with uncertainty: Information practices in the context of disease surveillance in Burkina Faso. *Information and Organization*, *31*(3), 100366. https://doi.org/10.1016/j.infoandorg.2021.100366.

Rawls, J. (1971). *A Theory of Justice: Original Edition*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9z6v.

Rawls, J. (1999). *A Theory of Justice: Revised Edition*. Harvard University Press.

Rawls, J. (2001). *The Law of Peoples.* Harvard University Press.

Rawls, J. (2005). *Political liberalism*. Columbia university press.

Ren, S., & Wierman, A. (2024, July 15). The Uneven Distribution of AI's Environmental Impacts. *Harvard Business Review*. https://hbr.org/2024/07/the-uneven-distribution-of-ais-environmental-impacts. Last accessed 25 Jan 2025.

Rességuier, A., & Rodrigues, R. (2020). *AI ethics should not remain toothless!* A call to bring back the teeth of ethics. *Big Data & Society*, *7*(2), 205395172094254. https://doi.org/10.1177/2053951720942541.

Rillig, M. C., Ågerstrand, M., Bi, M., Gould, K. A., & Sauerland, U. (2023). Risks and Benefits of Large Language Models for the Environment. *Environmental Science & Technology*, *57*(9), 3464–3466. https://doi.org/10.1021/acs.est.3c01106.

Risse, M. (2019). Human Rights and Artificial Intelligence: An Urgently Needed Agenda. *Human Rights Quarterly*, *41*(1), 1–16.

Romero Moreno, F. (2024). Generative AI and deepfakes: A human rights approach to tackling harmful content. *International Review of Law, Computers & Technology*, *0*(0), 1–30. https://doi.org/10.1080/13600869.2024.2324540.

Rowe, F., Ngwenyama, O., & Richet, J.-L. (2020). Contact-tracing apps and alienation in the age of COVID-19. *European Journal of Information Systems*, *29*(5), 545–562. https://doi.org/10.1080/0960085X.2020.1803155.

Sadek, M., Calvo, R. A., & Mougenot, C. (2023). Designing value-sensitive AI: A critical review and recommendations for socio-technical design processes. *AI and Ethics*. https://doi.org/10.1007/s43681-023-00373-7.

Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The sociotechnical axis of cohesion for the IS discipline: Its historical legacy and its continued relevance. *MIS Quarterly*, *43*(3), 695–720. https://doi.org/10.25300/MISQ/2019/13747.

Sarkkinen, J., & Karsten, H. (2005). Verbal and visual representations in task redesign: How different viewpoints enter into information systems design discussions. *Information Systems Journal*, *15*(3), 181–211. https://doi.org/10.1111/j.1365-2575.2005.00196.x.

Schiaffonati, V. (2022). Explorative Experiments: A Paradigm Shift to Deal with Severe Uncertainty in Autonomous Robotics. *Perspectives on Science*, *30*(2), 284–304. https://doi.org/10.1162/posc_a_00415.

Seger, E. (2022). In Defence of Principlism in AI Ethics and Governance. *Philosophy & Technology*, *35*(2), 45. https://doi.org/10.1007/s13347-022-00538-y.

Sen, A. (2010). *The Idea of Justice*. London: Penguin Books.

Siegmann, C., & Anderljung, M. (2022). The Brussels Effect and Artificial Intelligence: How EU regulation will impact the global AI market (arXiv:2208.12645). arXiv. https://doi.org/10.48550/arXiv.2208.12645.

Simons, J. (2023). *Algorithms for the People: Democracy in the Age of AI*. Princeton University Press.

Singer, A. (2015). There Is No Rawlsian Theory of Corporate Governance. *Business Ethics Quarterly*, *25*(1), 65–92. https://doi.org/10.1017/beq.2015.1.

Spil, T. A. M., Romijnders, V., Sundaram, D., Wickramasinghe, N., & Kijl, B. (2021). Are serious games too serious? Diffusion of wearable technologies and the creation of a diffusion of serious games model. *International Journal of Information Management*, *58*, 102202. https://doi.org/10.1016/j.ijinfomgt.2020.102202.

Spring, M. (2024, June 8). X removes accounts of network smearing politicians with deepfakes. *BBC*. https://www.bbc.com/news/articles/cq55gd8559eo. Last accessed 25 Jan 2025.

Stahl, B. C. (2007). Privacy and security as ideology. *IEEE Technology and Society Magazine*, *26*(1), 35–45. https://doi.org/10.1109/MTAS.2007.335570.

Stahl, B. C. (2008). The ethical nature of critical research in information systems. *Information Systems Journal*, *18*(2), 137–163. https://doi.org/10.1111/j.1365-2575.2007.00283.x.

Stahl, B. C. (2021a). Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies. Springer Nature.

Stahl, B. C. (2021b). AI Ecosystems for Human Flourishing: The Recommendations. In B. C. Stahl (Ed.), *Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies* (pp. 91–115). Springer International Publishing. https://doi.org/10.1007/978-3-030-69978-9_7.

Stahl, B. C. (2022). From computer ethics and the ethics of AI towards an ethics of digital ecosystems. *AI and Ethics*, *2*(1), 65–77. https://doi.org/10.1007/s43681-021-00080-1.

Stahl, B. C., Andreou, A., Brey, P., Hatzakis, T., Kirichenko, A., Macnish, K., Laulhé Shaelou, S., Patel, A., Ryan, M., & Wright, D. (2021). Artificial intelligence for human flourishing – Beyond principles for machine learning. *Journal of Business Research*, *124*, 374–388. https://doi.org/10.1016/j.jbusres.2020.11.030.

Stahl, B. C., Doherty, N. F., Shaw, M., & Janicke, H. (2014). Critical Theory as an Approach to the Ethics of Information Security. *Science and Engineering Ethics*, *20*(3), 675–699. https://doi.org/10.1007/s11948-013-9496-6.

Stahl, B. C., & Eke, D. (2024). The ethics of ChatGPT – Exploring the ethical issues of an emerging technology. *International Journal of Information Management*, *74*, 102700. https://doi.org/10.1016/j.ijinfomgt.2023.102700.

Stahl, B. C., Rodrigues, R., Santiago, N., & Macnish, K. (2022). A European Agency for Artificial Intelligence: Protecting fundamental rights and ethical values. *Computer Law & Security Review*, *45*, 105661. https://doi.org/10.1016/j.clsr.2022.105661.

Susskind, J. (2022). *The Digital Republic: On Freedom and Democracy in the 21st Century*. Bloomsbury.

Swanson, E. B., & Ramiller, N. C. (1997). The Organizing Vision in Information Systems Innovation. *Organization Science*, *8*(5), 458–474. https://doi.org/10.1287/orsc.8.5.458.

Tamássy, R., & Géring, Z. (2022). Rich variety of DA approaches applied in social media research: A systematic scoping review. *Discourse & Communication*, *16*(1), 93–109. https://doi.org/10.1177/17504813211043722.

Tomlinson, B., Black, R. W., Patterson, D. J., & Torrance, A. W. (2024). The carbon emissions of writing and illustrating are lower for AI than for humans. *Scientific Reports*, *14*(1), 3732. https://doi.org/10.1038/s41598-024-54271-x.

Umbrello, S., Capasso, M., Balistreri, M., Pirni, A., & Merenda, F. (2021). Value Sensitive Design to Achieve the UN SDGs with AI: A Case of Elderly Care Robots. *Minds and Machines*, *31*(3), 395–419. https://doi.org/10.1007/s11023-021-09561-y.

Umbrello, S., & van de Poel, I. (2021). Mapping value sensitive design onto AI for social good principles. *AI and Ethics*, *1*(3), 283–296. https://doi.org/10.1007/s43681-021-00038-3.

Vaidya, R. (2019). Corruption, Re-corruption and What Transpires in Between: The Case of a Government Officer in India. *Journal of Business Ethics*, *156*(3), 605–620. https://doi.org/10.1007/s10551-017-3612-5

Vakkuri, V., Kemell, K.-K., Jantunen, M., Halme, E., & Abrahamsson, P. (2021). ECCOLA — A method for implementing ethically aligned AI systems. *Journal of Systems and Software*, *182*, 111067. https://doi.org/10.1016/j.jss.2021.111067.

Vallor, S. (2016). *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford University Press.

Van Dijk, T. A. (2017). *Discourse and Power*. Bloomsbury.

Venable, J. R. (2010). Design Science Research Post Hevner et al.: Criteria, Standards, Guidelines, and Expectations. In R. Winter, J. L. Zhao, & S. Aier (Eds.), *Global Perspectives on Design Science Research* (Vol. 6105, pp. 109–123). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-13335-0_8.

Waelen, R. (2022). Why AI Ethics Is a Critical Theory. *Philosophy & Technology*, *35*(1). https://doi.org/10.1007/s13347-022-00507-5.

Wall, J., Stahl, B., & Salam, A. (2015). Critical Discourse Analysis as a Review Methodology: An Empirical Example. *Communications of the Association for Information Systems*, *37*(1). https://doi.org/10.17705/1CAIS.03711.

Walsham, G. (2012). Are We Making a Better World with Icts? Reflections on a Future Agenda for the IS Field. *Journal of Information Technology*, *27*(2), 87–93. https://doi.org/10.1057/jit.2012.4.

Weinhardt, C., Fegert, J., Hinz, O., & van der Aalst, W. M. P. (2024). Digital Democracy: A Wake-Up Call. *Business & Information Systems Engineering*, *66*(2), 127–134. https://doi.org/10.1007/s12599-024-00862-x.

Westerstrand, S., Westerstrand, R., & Koskinen, J. (2024). Talking existential risk into being: A Habermasian critical discourse perspective to AI hype. *AI and Ethics*. https://doi.org/10.1007/s43681-024-00464-z.

Westerstrand, S. (2024). Reconstructing AI Ethics Principles: A Rawlsian Approach. *Science and Engineering Ethics* 30(46), https://doi.org/10.1007/s11948-024-00507-y.

Wolkenstein, F. (2023). What is democratic backsliding? *Constellations*, *30*(3), 261–275. https://doi.org/10.1111/1467-8675.12627.

Young, A. G. (2018). Using ICT for social good: Cultural identity restoration through emancipatory pedagogy. *Information Systems Journal*, *28*(2), 340–358. https://doi.org/10.1111/isj.12142.

Young, A. G., Shuva, S., Roth, T., Zhu, Y., & Hevner, A. R. (2024). Ethical design through grounding and evaluation: The EDGE method for designing information systems for social impact. *Journal of Information Technology*, 02683962241289598. https://doi.org/10.1177/02683962241289598.

Zuboff, S. (2019). *The Age of Surveillance Capitalism*. London: Profile Books.

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama, Turku, Finland 2025

ISBN 978-952-02-0179-1 (PRINT)
ISBN 978-952-02-0180-7 (PDF)
ISSN 2343-3159 (Print)
ISSN 2343-3167 (Online)