DISSERTATIONES
SEMIOTICAE
UNIVERSITATIS
TARTUENSIS
46

AULI VIIDALEPP

The Expected AI as a Sociocultural Construct and its Impact on the Discourse on Technology

DISSERTATIONES SEMIOTICAE UNIVERSITATIS TARTUENSIS
46

DISSERTATIONES SEMIOTICAE UNIVERSITATIS TARTUENSIS

46

AULI VIIDALEPP

The Expected AI as a Sociocultural Construct and its Impact on the Discourse on Technology

UNIVERSITY OF TARTU
Press

1632

Department of Semiotics, Institute of Philosophy and Semiotics, University of Tartu, Estonia

The council of the Institute of Philosophy and Semiotics of the University of Tartu has, on September 25, 2023, accepted this dissertation for the defense for the Degree of Doctor of Philosophy (in Semiotics and Culture Studies).

Supervisor: Professor Timo Maran, PhD, University of Tartu

Opponents: Professor Kay Lunette O'Halloran, PhD, University of Liverpool, UK

Professor Donald Francis Favareau, PhD, National University of Singapore, Singapore

The thesis will be defended at the University of Tartu, Estonia, on December 18, 2023, at 15.00 in the University Senate Hall.

This research was supported by the University of Tartu ASTRA Project PER ASPERA (2014-2020.4.01.16- 0027), which is financed by the (European Union) European Regional Development Fund; European Social Fund's Doctoral Studies and Internationalisation Programme DoRa (Archimedes Foundation); the national scholarship programme Kristjan Jaak, which is funded and managed by Archimedes Foundation in collaboration with the Ministry of Education and Research; Estonian Research Council's institutional research project PRG314 "Semiotic fitting as a mechanism of biocultural diversity: instability and sustainability in novel environments"; Estonian Defence Forces' applied research project SSVUH20551 "Social aspects of applying uninhabited systems in framework of sociology of law"; European Commission's grant R-012 "Ethical, social and legal aspects of unmanned weapon systems"; EU COST Action CA19102 "Language in the Human-Machine Era"; Erasmus+ supports for study abroad (2022-1-EE01-KA131-HED-000054951). Erasmus+ is the EU Programme in the fields of education, training, youth, and sport for the period 2021-2027.

European Union
European Regional
Development Fund

Investing
in your future

ISSN 1406-6033 (print)
ISSN 2806-2582 (pdf)

ISBN 978-9916-27-415-6 (print)
ISBN 978-9916-27-416-3 (pdf)

University of Tartu Press
www.tyk.ee

To Eve,
who never lost faith in me.

ACKNOWLEDGEMENTS

This dissertation is the result of several years of work, during which I received help and support from many people and institutions inside and outside academia.

I am deeply grateful to my supervisor, Timo Maran, for his unwavering support and endless patience with my chaotic practices and unpredictable work capacity. His always insightful and thoughtful feedback ended up influencing my work implicitly much more than my initial topical focus would suggest. It is thanks to Timo's ecosemiotic and meta-scientific perspectives that I began to see the challenges of AI on a much broader global scale. Keeping ecosemiotics in mind, I eventually arrived at the Ecologies of AI conceptual model, which now underpins not only this work but is also instrumental for my future research.

I want to acknowledge the invaluable support of my colleagues and fellow researchers at the Department of Semiotics. I am deeply grateful to Ulvi Urm for her immense empathy and endless support. My heartfelt appreciation goes out to Kalevi Kull who always challenged me to push the boundaries of my research while staying true to semiotics. I very much appreciated the guidance of Mari-Liis Madisson during my master's studies and beyond. Her and Andreas Ventsel's classes on new media inspired me to start exploring the intersections of technology, society, and semiotics more closely, which further crystallized into my PhD topic. I extend a warm thank you to Daria Arkhipova for her constructive and rich feedback on my developing ideas, our shared ventures, and discussions. We will do great things together yet! I am also grateful to other Tartu colleagues who supported various stages of my studies and research: Peeter Torop, Silvi Salupere, Elin Sütiste, Maarja Ojamaa, Nelly Mäekivi, Katre Pärn, Silver Rattasepp, Ott Puumeister, Riin Magnus, Anti Randviir, and Tiit Remm. Jason Dydynski, thank you for editing my papers! Much gratitude to Merit Rickberg and Katarina Damčević for sharing many writing retreats and providing great inspiration on my PhD journey.

During my studies, I had many opportunities to travel and learn from other scholars in other universities. I am indebted to Barend van Heusden and Sjoerd-Jeroen Moenandar from the University of Groningen for their kind support in getting my PhD research started, and helpful feedback on my first paper. I would like to thank Massimo Leone who introduced me to his research group at the University of Turin. My deepest thanks go to Ľudmila Lacková who generously hosted me as a visiting student/scholar at the Department of General Linguistics, Palacký University Olomouc. She offered me complete academic freedom, trusting me with teaching and learning opportunities that allowed me to explore and expand my research topics and networks. I wrote most of my final draft in the lovely office at Palacký University which I will always miss! Other Tartu–Palacký colleagues Tyler Bennett, Israel Chávez, and Claudio Rodríguez helped me keep my mind sharp with various discussions.

Two projects have been essential for my PhD research. EU COST Action LITHME (Language in the Human-Machine Era), especially the working group on language ideologies and beliefs has offered me a much wider understanding of the significance of my work for other research contexts. Thank you to Britta Schneider, Bettina Migge, Theres Fester-Seeger, and others for involving me in their conversations. I am also deeply appreciative of the expertise and insights shared by my colleagues in the iMUGS project working group: Wolfgang Wagner, Kairi Talves, Eleri Lillemäe, Janar Pekarev, Markus Otsus. Our team discussions and studies of the public perceptions of AI greatly contributed to the outlining of my own research. I am equally grateful to everyone else I encountered throughout this project.

I am indebted to Maria-Kristiina Lotman and the team of Academia Artium Liberalium. Thank you for allowing me to explore my research in the form of public lectures for the Academia. This challenge brought so many new perspectives and discoveries that greatly enriched my work.

Many thanks to all students who participated in my various classes and challenged me to delve deeper and wider in my research topic, discover relevant research in other disciplines, and improve my expression skills. Without your inquisitive curiosity and intellectual contributions, this thesis would never have come to fruition. I am also grateful to fellow researchers encountered in conferences and seminars, online and offline. Their ideas and comments have been most inspiring. Especially the *Histories of AI* seminar series organised at the University of Cambridge in 2020–2021 introduced me to the works of many scholars who found their way among my favourite references.

A special thank you goes to everyone who has in some way joined or contributed to my *Technosemiotics* seminars, online and offline. Participating in a global interdisciplinary discussion is never easy—thank you for the curiosity and patience to join, listen, and ask (especially those outside semiotics!). Huge thank you to all the speakers for finding time and energy to present in this—still growing—discussion group. Your perspectives have challenged me to find ways to make semiotics more understandable and translatable for other disciplinary fields. I am truly grateful to Alec Kozicki and Josh Bacigalupi for joining me in this spontaneous endeavour and helping me chair the sessions. Josh has also been a steadfast friend and frequent lunch discussant on the topics of AI, capitalism, cybernetics, and biosemiotics.

My family has supported me throughout the toughest times, I could not have done it without them. My friends deserve equal gratitude. Eve saved me in more ways than one, and never gave up hope that I can do it, despite all the difficulties. My lifetime friend Elo-Mall restarted her university and research journey in the same year when I did. In these past years, we have shared more lecture-break-coffees and university discussions than I can remember. Marge joined my worries about the impact of generative AI on artists' lives and work. Kaur provided me with a steady supply of AI novelties and use cases (I suspect several of those have made it to this thesis). Ehte offered me frequent company in Tartu cafés and our self-made writing retreats. She also occasionally lent me Geko, the little four-legged semiosic creature who is the perfect thesis and therapy dog. Geko patiently oversaw various phases of my writing and research. A special thought to Luna, Anu, Savannah, Sven, Oscar, Jin, Eleni, and Andrew in continuing and finishing your own endeavours. And finally, the biggest hug to Lona and the seals! For all the conversations we've had, writing together, despairing together, and finding our motivations again. *Bon courage!*

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

I. Viidalepp, Auli 2020. Representations of Robots in Science Fiction Film Narratives as Signifiers of Human Identity. *Információs Társadalom* 20 (4): 19–36. doi: 10.22503/inftars.XX.2020.4.2.

II. Viidalepp, Auli 2022. Sociocommunicative Functions of a Generative Text: The Case of GPT-3. *Lexia. Rivista Di Semiotica* 39–40: 177–92. doi: 10.53136/979122180426310.

III. Viidalepp, Auli 2022. The Semiotic Functioning of Synthetic Media. *Információs Társadalom* 22 (4): 109–18. doi: 10.22503/inftars.XXII.2022.4.9.

The purpose of a thought-experiment,
as the term was used by Schrodinger and other physicists,
is not to predict the future—
indeed Schrodinger's most famous thought-experiment
goes to show that the 'future,' on the quantum level, cannot be predicted—
but to describe reality, the present world.

Science fiction is not predictive; it is descriptive.

— Ursula K. Le Guin, Introduction, The Left Hand of Darkness

INTRODUCTION

In 1976, Ursula Le Guin published the second edition of her space novel *The Left Hand of Darkness*. Compared to the first edition, the second print had gained an introduction. In this introduction Le Guin lays out what she sees as the function of science fiction or literature in general—to be a thought experiment, reflective of its contemporary society. She stresses multiple times that science fiction is not about the future, and it should not be read as a prediction.

Computers are sociotechnical as well as semiotic systems. However, media discourse as well as marketing discourse from technologists themselves tends to conceal the human labour necessary for AI systems to be deployed and for them to continue functioning. But the question is not necessarily about when and how AI will "take our jobs". Instead, I ask: how do we arrive from hundreds, thousands or even millions of people working for several years to achieve something like the system of ChatGPT, released into the public, to the discursive portrayal of said system as a single entity actant on the verge of "becoming sentient"?

The main purpose of this work is to introduce and provide some critique on the discourse on technology, with a specific reference to the concept of artificial intelligence (AI). The discourse on AI is particularly saturated with reified metaphors which drive connotations and delimit understandings of technology in society. This is true for both academic and public media discourse.

My premises and arguments set out for exploration in this thesis are the following:

- The discourse on AI is entangled with (science) fiction, folklore, myth, religion. This entanglement impacts how AI is perceived and received, as well as the expectations of AI-enabled technologies now and in the future.
- The cultural-historical roots of AI as a concept and as an idea contribute to the formation of an "Expected AI". This composite signifier is filled and fed by historical and sociocultural connotations, various referent objects etc.
- The discourse on AI is strongly influenced by the anthropomorphism of technology, which forms another facet of technological determinism or the idea that technology develops without human interference. Many technologists overtly or covertly support and perpetuate this idea.

To reach these aims, I proceed as follows.

Chapter 1 addresses the historical entanglements between semiotics and AI (1.1; 1.1.1) and summarises the main ways that semiotic metalanguage has been used to address the ontological questions of AI and computer systems more generally (1.1.2).

In Chapter 2 I give an overview of the "Expected AI" as a composite signifier (2), its history (2.1) and contemporary connotata (2.1.2). I reveal the cultural repositories this complex idea retrieves its meanings from, which include science fiction, folklore, and various contemporary philosophical ideas. Section 2.1.1

frames AI as a political device, with political actors seeking to employ the power of the miraculous to cement their power via sociotechnical imaginaries and technological developments. The section further connects the idea of AI as a political device to similar tendencies from Ancient Greece.

Subchapter 2.2 addresses various problems with metalanguage in AI discourse such as discursive anthropomorphism of technology (2.2.1) and its serious consequences (such as automation bias, 2.2.2). Subchapter 2.3 asks whether we can talk about technology as neutral, and who—in current politics—is allowed to talk about technology at all.

I am also writing this work in the hope that by pointing out the sociocultural roots of the various anthropomorphic connotations of AI, we can move beyond anthropomorphism and technological determinism and ask better in what ways the technologies are being constructed and manifest as they are, and what kind of impact, problems, and opportunities they thus bring to all the people involved right now. To this end, Chapter 2.4 proposes a more ethical and comprehensive model for AI systems, considering their complex socio-material organisation, global economic-material becoming, and impact. This chapter outlines an initial idea for a heuristic for understanding AI as a complex system that also accounts for environment, social structures, institutions, and the semiosphere.

Overall, I describe AI systems as any other computer systems—inherently semiotic and comprised of various human actors as well as techno-material elements. I reveal the discursively concealed human roles within these systems and seek directions of sociotechnical developments that would further empower the people working in and impacted by AI in the world today.

It has been argued that AI is not a science, but an engineering practice. However, many AI builders (and researchers) see themselves as scientists, reflecting the composition of cultural practices and building applied technological tools that act upon cultures and organise or influence these practices, people, societies and so on. Despite arguing to carry and implement the scientific perspective, they often do not comply with the good and critical scientific practices while doing their work. On the contrary, they resurrect buried pseudo-scientific theories (e.g. phrenology!) and propose implementing these in decision-making systems that act upon real human communities.

When I started my research journey into AI—first, by curiously following the increasing media discourse in 2016—it began from a similar premise, informed by science fiction and the general debates on the media, that our technological developments are, slowly but surely, moving towards some kind of "Artificial General Intelligence" (AGI). Then, starting the PhD project in 2019 and reading further and further into various fields of research, including Science and Technology Studies (STS), philosophy of mind, cognitive science, and semiotics, I very quickly realised that it is not that simple. Based on newly acquired information and latest research, I gathered that AGI was mostly a myth-based narrative that was, for some reason, extremely popular. So, I set out to research that myth with my PhD. Throughout my journey, I learned a few things: 1) that in some ways the idea of AGI is not as wide-spread and cemented as I feared; and 2) that there are many other researchers that are doing pretty much what I am, by looking at ways to more adequately describe and analyse what is happening at the intersection of AI, politics and society; and 3) that the general confusion about AI/AGI and related ideas is quite persistent but not that loud.

In the intermittent years, the discourse and illustrations seemed to come a long way from the fiery-eyed Terminator pictures I saw accompanying every other article about AI back in 2018. With the applications of concrete tasks on concrete devices (such as vacuum bots, or even military technology) it seemed like it was possible to remain in the present world while talking about AI.

Then in 2023 everything changed.

A pertinent example of how well spread the confusion and misunderstanding about AI harms is, is illustrated by the recent AI embargo letter.[1] Its signatories demonstrate their gullibility to the eschatological myth, while remaining ignorant of the real harmful impact of current AI systems on societies, cultures and above all, climate.

Also, the subsequent tweet-like statement[2] proclaiming AI as an existential risk belies a lot. It again raises AI to the level of the nuclear (a category mistake); and it completely disregards our current, most pressing existential risk posed by climate change and environmental issues. Admittedly, some leaders in AI have recognised this issue and professed their wish to "develop an AGI that would solve the climate problem". However, as Naomi Klein[3] argues, the main issue around climate change is not that we lack scientific models or scientific understanding of the problem and its possible solutions or alleviations. The lack of solution is not scientific but political: there has been for some time a scientific consensus for what needs to be done; it is simply the political leaders in the world that fail to instigate measures to move towards these solutions. And this is not an issue that can be solved with developing more computer programs.

Or would the issue perhaps be more easily solved—and taken more seriously by the leaders—if, instead, a *machine* suggested the same solutions that scientists have been talking about for decades? Such an occasion would, indeed, be an ironic twist in the realisation of the extent of anthropomorphism and automation bias in our societies.

---

1 "Pause Giant AI Experiments: An Open Letter." n.d. *Future of Life Institute* (blog). https://futureoflife.org/open-letter/pause-giant-ai-experiments/. Accessed September 8, 2023.

2 "Statement on AI Risk." n.d. *Center for AI Safety*. https://www.safe.ai/statement-on-ai-risk. Accessed September 8, 2023.

3 Naomi Klein, "AI Machines Aren't 'Hallucinating'. But Their Makers Are." *The Guardian*, May 8, 2023, https://www.theguardian.com/commentisfree/2023/may/08/ai-machines-hallucinating- naomi-klein. Accessed September 8, 2023.

Summaries of the articles

(I) Viidalepp, Auli 2020. Representations of Robots in Science Fiction Film Narratives as Signifiers of Human Identity. *Információs Társadalom* 20 (4): 19.

The article analyses robot characters in four recent science fiction series. Employing the concepts of semiosis, unpredictability, and art as a modelling system, as well as Paul Ricoeur's concept of narrative identity, the article frames the robot characters as signifiers of extratextual human Other, thus revealing the films' narratives as typical literary and thought experiments with human identity. Additionally, utilising Roslynn Haynes' scientist stereotypes, the research recontextualises the characters of robot makers as instances of the "crazy scientist" stereotype. The article also tentatively outlines the fictional superhuman spectrum, consisting of all characters that are superhuman, supernatural, cyborgs, monsters, or artificial creatures.

(II) Viidalepp, Auli 2022. Sociocommunicative Functions of a Generative Text: The Case of GPT-3. *Lexia. Rivista Di Semiotica* 39–40: 177–92.

The article takes into focus the first public and highly publicized experiment with generative AI—the case where *The Guardian* editors used GPT-3 by OpenAI to generate eight texts on the topic of why humans should not fear AI. The eight results were then partially combined and published as an "opinion article". Consequently, several new media outlets posted commenting articles where several chose to quote the op-ed as if it were the "robot's opinion." The publication and reception of the op-ed highlights the difficulty for human readers to differentiate a machine-produced text; it also calls attention to the challenge of perceiving such a text as a synthetic text even when its origins are made explicit. The article offers a critical examination of the process behind the generation and the interpretation of a synthetic text, using Lotman's concept of text and its sociocommunicative functions. The article analyses how people are inclined to perceive a relatively well-written text as an intentional message, and how generative media conceals and displaces the habitual author in relation to texts.

(III) Viidalepp, Auli 2022. The Semiotic Functioning of Synthetic Media. *Információs Társadalom* 22 (4): 109–18.

This article takes a more general and theoretical view on generative media, focusing on the generated text and deepfake video technology. The article outlines certain problems with AI-generated content and frames the issue as a problem of recognising its proper referential reality. The article makes use of the concepts of mimicry and nonsense to reveal the elements and counterparts in the communicative processes involving generated media. The article also discusses the referential shift occurring with synthetic media, and some general principles of its semiotic functioning.

1. METHODOLOGY AND DISCIPLINARY BACKGROUND

When it comes to the interpretations of AI in society and public media space, the real issue at hand is the formation of the public understanding of science – how the public knowledge of AI, specifically, is constructed and what kinds of repositories and relevance structures it draws upon. The dominant models for this—the diffusionist model and the deficit model – cannot successfully explain the intricacies of public knowledge (Sismondo 2010: 170–74). The diffusionist model assumes that scientific knowledge is too complicated to be widely understood. The deficiency model assumes that scientific literacy is the problem of education or lack thereof (ibid.). Both models lack explanatory power for how the belief in conspiracy theories and recognisably pseudoscientific views, for example, is unexpectedly persistent in countries that are fully literate, have accessible and scientifically sound education systems, and should therefore have all the prerequisites for holding a scientifically literate population.

The question of AI, its meaning, impact, and role in society is the most related to science and technology studies (STS). STS is a relatively new academic field and the latest paradigm addressing the main questions discussed in the philosophy of science and the sociology of knowledge, joining them with the general problem of the impact of technology on science and society. The ontology and nature of scientific knowledge and its distinction from other kinds of knowledge has long been debated in the philosophy of science, formulated in terms of scientific paradigms (Popper), scientific revolutions (Kuhn), or demarcation lines (Lakatos). The focus of all these theories is on how some knowledge becomes defined as scientific and a valid description of reality, and how to recognise it as such from its representational structure. Usually, these approaches do not account for the sociocultural dimension in the production of scientific knowledge and look for essentialist definitions of science. STS, instead, takes the anti-essentialist premise that the production of knowledge in science and technology is thoroughly social; that neither field is in any way 'natural' or pre-given, there being no guaranteed method that can "translate nature into knowledge" or "translate knowledge into artifacts"; and that there always remain multiple interpretational possibilities when it comes to knowledge and artifacts (Sismondo 2010: 10–11). Thus, the field of STS inquires how knowledge and artifacts are *socially constructed*. Sismondo (2010) explains this mostly in terms of actor-network theory (ANT) (Latour and Woolgar 1986) but contends that ANT has certain limitations and cannot fully account for the social side of the networks at play.

ANT and similar structural models tend to treat the sociocultural reality as assemblages where all the elements (people, objects, institutions, and the relationships between all these) are treated in a way that obscures human agency within those assemblages. The model certainly allows for observation and discovery of relationships and their impact on sociocultural reality regardless of the activity or passivity of the individuals. However, the treatment of objects and people as elements of the same flat ontology may leave a deceitful impression that human agency is not at all necessary when analysing technological developments. Therefore, such models need to be used critically and with caution, to avoid unnecessary technological determinism.

The object of research for this thesis is composed of sociocultural representations of AI, or various cultural texts in a Lotmanian sense. I recognise the signifier "AI" as an empty signifier (Lévi-Strauss, Laclau) which—due to its emptying of a single, fixed, distinct denotation—tends to attach itself and be attached to a plethora of connotations that originate from different points in the semiosphere, forming a more or less coherent descriptive system characterised as the Expected AI. As I explain in various subsections of Chapter 2, the connotations for AI are borrowed from various discourses that are occasionally contradictory and quite different from one another. As a composite signifier for hope, fear, humanity, identity, and technology, AI tends to be associated with various historical-cultural texts and intertexts that discuss these themes. Such connotative topics include various myths, mythical and literary characters; mechanical objects and stories thereof; and almost the entire genre of science fiction. The intertexts date back to Ancient Greece at least, and fictional, non-fictional, public, and academic discourses alike contribute to remixing, reusing, and perpetuating these associations.

1.1 AI and semiotics

"Words can be used thus paradoxically because they have, along with a semiotic usage, a symbolic or metaphoric usage."

— Ursula K. Le Guin, The Left Hand of Darkness

Semiotics can and has been used to address the issues of AI on at least three separate levels: to analyse the discourse, to analyse the ontology, and in leveraging semiotics for building better AI models. While AI has a long-standing common history with philosophy within the paradigm of the philosophy of mind, the dialogues and common points between AI and semiotics are not so well known.

For Juri Lotman, semiotics was already an amalgamation originating from the intersection of scientific fields prominent in the 1950s such as structural linguistics, information theory, cybernetics, and logic (Salupere 2015: 69). Incidentally, all those disciplines also underlie most of AI as a field. With so much in common, why are semiotics and AI not more intertwined today?

1.1.1 *Machine semeiology* and the historical Dartmouth 1956 AI camp

It is notable that semiotics, or the "science of signs", was brought up in the historical Dartmouth camp by then-student Trenchard More. More's participation proposal for Dartmouth, titled "Computer decisions in deductive logic" (More 1956a) does not make explicit this connection, nor does it include any remotely semiotic references, save for mentioning “the natural deductive systems of Quine and Fitch” on the cover. His four-page proposal rather focuses on explaining how digital computers’ ability to process arithmetic data could be extended to manipulating data with “symbolic significance”—that is, language, or objectivations of it—by employing certain metamathematical concepts proposed by Gödel, Kleene, Church, and Turing. More repeatedly emphasizes that the proposed system does not employ any decision procedure within the machine. Instead, he envisions “to have the computer learn pertinent theorems from its operator. In this way, the capricious and selective elements so necessary to intelligence are supplied from a human source” (More 1956a: 4). The stated goal of the project, in line with the general topics of the Dartmouth camp, is to “direct a machine to solve problems related to language translation, inventory control and business decisions” (More 1956a: 2).

In the report on the two weeks that More spent on campus, where he summarizes the talks and discussions for his intern employer IBM, he explicitly states the connections to semiotics. According to the report, relying on the theories of Frege, Quine, Peirce, “Morris’s behavioral semiotic,” and Carnap, among others, More proposes a direction of research for formulating a *machine semeiology*, which would be “the science of signs applied to the design of machines, as well as the study of machines applied to the understanding of signs” (More 1956b: 5). More’s (1956a) proposal departs from Gödel’s incompleteness theorem, suggesting the idea of the extra-systemic decision element (the operator). By the end of the camp, as per the report, More finds “that his program is guessing in a metalanguage, rather than in the object language,” leading him to search “for a way to handle several levels of language at once” (More 1956b: 8).

However, his ideas were not unilaterally well received at the conference.

> “One thing that this study project has made clear: that Minsky, MacCarthy, Selfridge and myself are all aiming for the same goal. Ashby is, too, but he would, at present, be satisfied with more modest returns. Simon may be serious about the ultimate goal—and perhaps Newell. Shannon doesn’t yet see anything worth working on that is promising. With Trench More, its (*sic*) hard to say. He will change—he is probably not too hopeful yet, but probably has picked up much from the spirit of the group—apparently More has come from rather negative surroundings, and they (and other surroundings) tend to influence him strongly.” (Solomonoff 1956)

Ray Solomonoff in his handwritten notes takes issue with More’s attitude, considering it too pessimistic regarding the future of “thinking machines” (Solomonoff’s own notes are full of references to “TM”—“thinking machine”); he attributes it to More coming “from negative surroundings”, whatever that means, but expresses hope that More will change.

There are no further references to Peirce or Morris in the few articles More subsequently published. Instead, More went on to develop Array Theory, a mathematical system of nested arrays that provided the logical foundation for the construction of several programming languages later on (More 1973).

1.1.2 Using semiotics to define the ontologies of AI

Semiotics is sometimes used to define the ontologies of AI. Theoretical biology does not offer any further explanations about why and how some things are alive and others are not; so biosemiotics has taken up this debate and offers answers grounded in Peircean semiotics, cybernetics and systems theory. Further, the same delimitations are applied to the discussions about the potentiality of machine consciousness. This debate can take at least two directions: one is focused on the problem of where, when, and how complex systems manifest living properties (ergo the question: could machines manifest similar properties when they become complex enough?); the other debate is about whether and to what extent humans and other living beings are describable and/or heuristically analysable by computational models. The epistemological value of computational models and paradigms is evident—computational neuroscience and other fields have helped make great advancements in understanding the functioning of human and other living organisms.

Semiotics is often defined as the science about communication in living organisms. The processes of communication and signification make up *semiosis*, "an indispensable characteristic of all terrestrial life forms" (Sebeok 1991: 22). Sebeok treats semiosis "as the defining feature that criterially distinguishes the vital from the inanimate" (Sebeok 1986: 10). In this, he follows the Umwelt-theory of Jakob von Uexküll, who also "viewed semiosis as the criterial attribute of life" (Sebeok 1979: x), even while he did not use the exact same word. Sebeok also emphasizes that "Semiosis on a superior level in the hierarchy of integrons is irreducible to that on a lower level, namely, ultimately to physics" (Sebeok 1991: 14)[4].

Considering the above, it is somewhat surprising that in the *Handbook of Semiotics* series (Posner et al. 1997, 1998, 2003, 2004), edited by Roland Posner, Klaus Robering and Sebeok himself, the concept of *semiosis* has made it into a configuration of *machine semiosis.* In the course of extending the theory of semiosis from human-centred anthroposemiosis to other living processes, and perhaps due to the renewed AI frenzy of the late 1990s, the concept of *machine semiosis* is also considered (Andersen et al. 1997). According to the chapter, "'Machine semiosis' denotes the semiotic processes that take place inside machines, *between machines*, and between them and their human users." (Andersen et al. 1997: 548, added emphasis) The authors first consider machines as *representamens* (ibid: 549–552), departing from Eco's definition of the sign as "everything which can be used in order to lie" (Eco 1976: 7) and Rossi-Landi's typology of artifacts, in which 'automated machines' form the highest order. They then turn to machines as *interpretants* (Andersen et al. 1997: 552–555), although they seem to conflate the

[4] In the second edition of the same book (Sebeok 2001: 29), this quote has, for some reason, changed from "irreducible" to "reducible" which, I am sure, is a typographical error.

concept with that of interpreter[5] and focus rather on how humans colloquially anthropomorphize computers and simply tend to attribute semiosis to computers based on their own language use (ibid: 554). The authors suggest this is because, while ordinary signs can create social referents, they cannot create physical referents, but computer-based signs can. This situation causes the computational process to appear somewhat "magical", therefore it is "no wonder we sometimes confuse the map with the landscape" (Andersen et al. 1997: 555). Indeed, "machine learning both *represents* and *intervenes*" (Roberge and Castelle 2021: 13, original emphasis) with our understanding of the external world, and one of the most common complaints about AI systems is against their biased filtering of our world experience and our uncritical acceptance of that filter.

As the authors conclude, the idea of a true semiosis happening in a machine is out of question, therefore "machine semiosis" inevitably presupposes human semiosis (Andersen et al. 1997: 569). Eventually, we are left with the idea of basic human-computer interaction, or humans using the computer as a tool, in other words—a kind of database query that gives some results. Leaving aside the detailed signifier for the process happening inside the computer, it becomes questionable why a separate term is needed at all to describe the sign process happening in a human working with a specific tool—even if it is part of somewhat novel "intelligent technologies". The situation reminds of Peirce's famous "inkstand argument":

> "A psychologist cuts out a lobe of my brain [...] and then, when I find I cannot express myself, he says, 'You see, your faculty of language was localized in that lobe.' No doubt it was; and so, if he had filched my inkstand, I should not have been able to continue my discussion until I had got another. Yea, the very thoughts would not come to me. So my faculty of discussion is equally localized in my inkstand. It is localization in a sense in which a thing may be in two places at once." (CP 7.366; cited in Skagestad 1999: 552; Nöth 2002: 90)

Winfried Nöth displays the very quote in his 2002 article *Semiotic Machines*, followed by an argument for the *quasi-mind* located equally in the extended medium, as well as in the human brain. Nöth seems to disregard two issues that I consider essential for this context in the Peircean theory. First, the above passage about the inkstand is a dripping irony aimed to criticize the reductionist thinking of Peirce's contemporary psychology (that reduced language to an area in the brain). Second, Peirce's usage of the *quasi*-preposition in various contexts (quasi-sign, quasi-mind etc) is not necessarily intended as a distinct term but rather taken for its colloquial Latin meaning (seemingly, as if).

If we take the "inkstand semiosis" too seriously, we can claim in a similar fashion a whole typology of various "semioses" wherever a human is using a certain device to extend their physical or cognitive abilities—such as "paper semiosis", "typewriter semiosis", "letterpress semiosis", "pencil semiosis" or

---

[5] "We have already seen that tools and machines act as signs for their human users: but how should we tackle machines that seem to respond to signs? Should we place machines in the interpreter role?" (Andersen et al. 1997: 552)

"knife semiosis". The flexibility of language and semiotic terminology certainly allows for such distinctions. However, it quickly appears ridiculous and unnecessary to specify such details, because the semiosis happening everywhere is the same—taking place in the human involved—and even Peirce's jested advocacy for the agency and importance of his inkpot in formulating his thoughts does not actually justify why such a metalinguistic construct is needed. This is especially true when it would serve to further confuse the boundary between human and machine and increase the chance for an anthropomorphic fallacy, which is problematic in human–computer interaction (HCI).

Arguments against reductionism have since appeared in other semiotic discussions, sometimes in polemics with the "Dartmouth AI". Mihai Nadin especially emphasizes that "contrary to statements made since Dartmouth, human beings are not reducible to algorithmic machines, and a science of the human being transcends the algorithmic description" (Nadin 2019: 217).

Certain general semiotic theories and concepts seem to have been developed in response to the emerging field of AI. Michael Polanyi's formulation of *tacit knowing* (Polanyi 1958, 1966) was, in part, an attempt to make explicit how and why certain kinds of knowledge are not reducible to computational models. Umberto Eco's theory of *s-codes* outlines in a similar manner the distinction between signal-based and symbolic interaction (Eco 1976). For Doede Nauta, the "automaton" type of system corresponds to signal semiosis as well (Nauta 1972: 131).

Charles Peirce has inspired a distinct strand of thinking based on his mentions of quasi-signs and his understanding of mind as something partially externalised to the human body via 'mind-enhancing tools' such as pen, paper, and inkstands (Peirce 1887; Jorna et al. 1993; Skagestad 1993, 1999; Emmeche 2007; Nöth 2019). Irvine, however, argues that Peirce is the rightful progenitor of information theory entirely, having developed his ideas of 'differential calculus', Logic Machines and Existential Graphs in the first decade of the 20th century:

> "Peirce clearly understood how binary logic maps onto switched electrical circuits, and how a binary mathematical code could be used for electrical signals, but these applied semiotic ideas had to wait for their application in the 1930s, when rediscovered by Claude Shannon for telecommunication networks and binary data." (Irvine 2022: 210)

Following Peircean thought, some authors tentatively continue blurring the boundaries between semiotics and machines (Nöth 2001, 2002, 2008; Nadin 2007, 2011; Colapietro et al. 2020). The quasi-sign doctrine pertinent to dyadic signs (including the ontologies of machine-mediated communication) is now being developed further by Tyler Bennett (2021). The quasi-sign doctrine remains out of the scope of this thesis. However, further research on quasi-sign or dyadic ontologies is necessary and urgent now, to find ways to cope with the proliferation of synthetic, machine-generated content on the Internet.

Elsewhere, biosemiotician Donald Favareau argues that the whole ensemble of complexity sciences—systems theory and its scions cybernetics, catastrophe, and chaos theory—can be traced back to Uexküllian *Funktionskreis* (Favareau 2010: 42). Using Uexküll's somewhat cybernetic vocabulary to discuss the ontologies of computer and AI systems is yet another strand of theorizing which remains out of scope for this thesis.

In my opinion, the problem lies partly in the uncritical use of the concept of 'semiosis' regarding machines—a practice by computer scientists and semioticians alike. For example, computer scientist Erich Prem (1998) talks about the process of semiosis in robots as 'autonomous sign users', employing Uexküll's functional circle as a model for better construction of robot's input-output processing. He does not make an explicit difference between signals and symbols, and although he refers to Peirce, he does not touch upon the concept of quasi-sign as Nöth does. From a robotics development point of view, it does not make much difference that computer-based models find inspiration in models of organismic sign processing. They have followed concepts of animal behaviour already (Ziemke and Sharkey 2001 mention this). However, it is more problematic when computer scientists, relying on Uexküll's theory insist that robot creatures are just like animals (by claiming both have 'species', see Ziemke and Sharkey (2001: 716) who criticize the metalanguage of Rodney Brooks), or possess first-hand semantics (ibid: 703) or have a subjective view of the world (ibid: 739). AI researcher Erich Prem (1998) refers to both animals and robots as 'embodied autonomous systems' (Ziemke and Sharkey 2001: 703).

Ziemke and Sharkey (2001: 722) talk about a "semiosis of information" in a robot built on an artificial neural network (ANN) "where the input corresponds to the sign, the input-output mapping to the interpretant (or causal rule), and the output to the signified". They mention "computer programs which are to some degree also capable of semiosis" (2001: 730), relying on Andersen et al (1997) with this claim. The "to some degree" semiosis in this case is, then, more the kind of dyadic or "signal semiosis" (Nauta 1972).

Jean-Guy Meunier (1989) sees AI as applied semiotics. Meunier (2021) discusses the possibilities of computational semiotics. He emphasizes the critique on and serious shortcomings of computational models in semiotics, which can be "too 'distant' from their true object," leading to superficial and uncritical analyses. However, he stresses that these criticisms do not oppose computers and computing as such but are merely critical of the discursive overselling and excessive trust placed in such models. The models can be found completely acceptable if we recognize that "[a]s models, they are by definition partial and reductive," and "cannot deal with non-computable relations and functions." Indeed, the real "danger is to believe that they can." (Meunier 2021: 208)

The most recent overview comes from Martin Irvine (2022) who develops a theory of computer systems as inherently semiotic assemblages, counting on the people (humans) whose intentions and agencies are intertwined in a complex manner with the sociotechnical elements, thus producing new types of interactions where recognising any individual intention vector is not that simple any more.

> "The 'physical symbol system' descriptions get us part way to a semiotic model, but the theory is based on an impoverished conception of signs and symbols, mostly modelled on the formal symbols of symbolic logic notation, with rules for logical operations and relations, that can be assigned in the computer architecture. The role of semiotic agency and the function of interpretant relations in a full triadic symbol system model, in Peirce's sense, are unaccounted for. However, the assumptions and terminology of the 'physical symbol system' hypothesis continue to inform arguments about symbols in theories of computation, cognition and AI […]." (Irvine 2022: 221–22)

Revisiting semiotic theories is particularly relevant to the increasing popularity of computational language manipulation models. Natural language processing (NLP) has been a challenge as well as a success in AI development since the very beginning. Latest advances in machine learning have led to the proliferation of different language models (LM), sometimes also called large language models (LLM) due to the high number of parameters and large amount of data used for their training.

The quote from Martin Irvine serves a pertinent introduction to the following chapter.

1.2 Computer systems are inherently semiotic and socially constructed

All computer and information technologies are, in fact, essentially and inevitably semiotic exactly because they manipulate symbols: "Any computer system, large, small or unobservable, represents an implementable design of applied semiotic structures in a unified architecture based on, and in the service of, human symbolic thought." (Irvine 2022: 224) However, this relationship is better characterised as an analogue to the relationship between a map and the terrain: "Computer system design principles provide homological maps for symbolic to physical correspondences that enact assigned representations and operations." (Irvine 2022: 224)

> "From a pragmatist semiotic perspective, the […] 'computer system' is actually the *whole dialogic supersystem* comprised of semiotic agents (aka 'users'), who are not independent individuals but members of meaning-making communities, and computer *systems* embodying semiotic system design for dialogic interaction." (Irvine 2022: 224, original emphasis)

Computer systems, including any implementations of AI, consist of a large and complex network of relationships between hardware, software, and foremost humans who are present and envisioned within the large system as planners, designers, programmers, system architects, project managers, product marketers, and eventual users of the given products. This composition makes computers thoroughly social and sociocultural constructs.

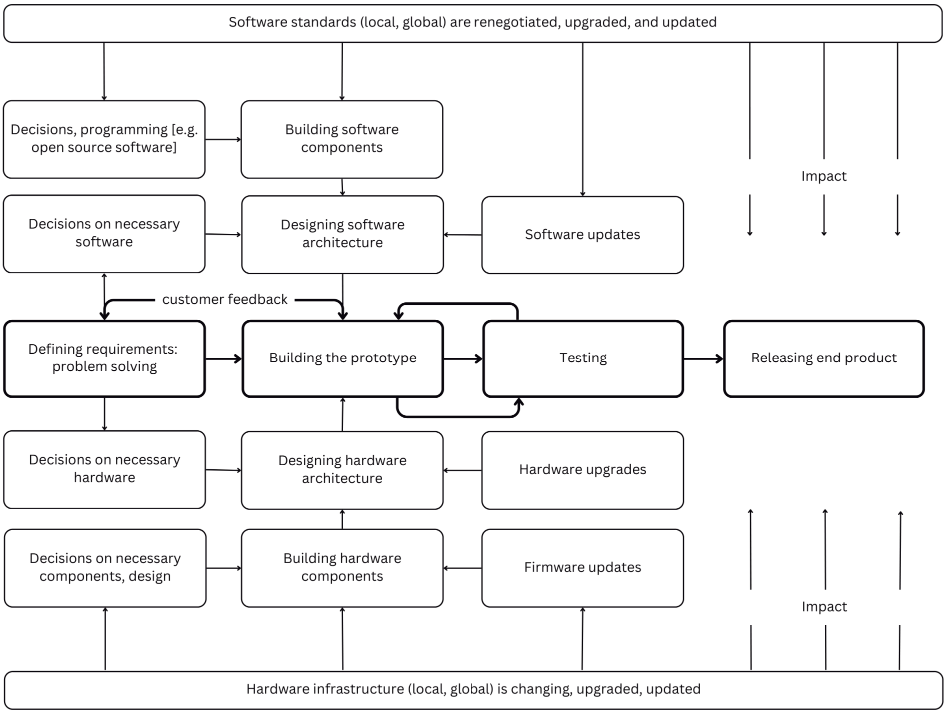

Figure 1. A diagram depicting the social construction of a computer product. Every box on the diagram represents the choices and labour of many people. The core design process is also influenced by various external activities reflecting the ever-changing infrastructure and industrial standards.

As shown in Figure 1, every small detail in the long and complex product cycle of an AI system is the direct or indirect result of choices made by humans within these assemblages—from the shape, size, and colour of the laptop to the design, appearance, and behaviour of every single software and hardware component. Every instance, component, and function of a computer system is the result of creative work by its programmers and engineers, combined with the feedback of the customer/client, and continuously adapted to the changing constraints of the wider techno-social environment, characterised by the developing infrastructure and software standards. One example where this creativity comes to the fore is the constant problem with systems interoperability. All systems are constantly developed further and at some point, they stop collaborating with older components because the developments have become too incompatible. To alleviate this issue, constant software and firmware updates are developed, which requires human work. Unfortunately, many technology producers have resolved to provide updates only to a certain point, after which the end user is forced to change the entire device – discard the old and buy a new one. The specifications of standards—both in industry and in the web—is subject to negotiations in specific institutionalized bodies such as W3C (World Wide Web consortium) or IEEE (the Institute of Electrical and Electronics Engineers). Considering the deeply human composition of all these institutions, and the enormous work and social negotiating spent in that process it is a wonder that any technological specialist would ever consider computers as deterministic, completely independent of human will and input.

The Club of Rome 1972 report *The Limits of Growth* discusses iterations of a world model estimating the constraints to economic growth posed by various resources. Even while the authors try to consider the impact of several technologies (such as birth control pills, television, or offshore drilling), they admit that these impacts are too unpredictable for the model. While enabling to an extent estimates of the economic and physical side effects of a given technology, "Unfortunately the model does not indicate, at this stage, the social side-effects of new technologies. These effects are often the most important in terms of the influence of a technology on people's lives." (Meadows et al. 1972: 146)

1.2.1 The representationalist fallacy

Irvine's analogue of symbolic mapping (Irvine 2022: 224) echoes an earlier critique from the authors of the *Machine semiosis* chapter about confusing the map with the landscape (Andersen et al. 1997: 555) when it comes to the relationship between the symbols and the reality, or Philip E. Agre's critique of the "conflation between representations and [various] aspects of AI research" (Agre 1997a: 143).

Agre criticizes the way AI researchers disregard linguistic referential precision when attributing inherently anthropomorphic vocabulary to describe their models' functioning, to the extent that the vocabulary of computational models is conflated with the ideas about external reality. This leads to inherent conceptions of reality that are confined to the computational model. "It is common for AI people (and computer scientists in general) to employ the same words to name both the representations in a machine and the things that those representations represent," argues Agre (2002: 133–34). As a result, the representations (signifiers) become conflated with their objects (signifieds), to say nothing of flattening the entire triadic sign structure (representamen-interpretant-object in Peirce's terms). This removes the possibility for self-critical analysis:

> "The point, therefore, is not that AI researchers believe that the mind and world are the same, or that they are incapable of grasping the idea that they are different. Rather, practices that tend to conflate the two are dispersed throughout the discourse and practices of the field; researchers are socialized into the discourse and practices through their training in the field, and the field does not provide the critical tools that would be necessary to detect this pattern and reverse it." (Agre 2002: 134).

One of the side effects of representationalist thinking is the discursive reification of technology. As "technical languages encode a cultural project of their own (the systematic redescription of human and natural phenomena within the limited repertoire of technical schemata that facilitate rational control)" (Masís 2014: 61), they quite literally make things with words then, as Austin said (Austin 1962).

For Agre, AI is also a discursive practice with a discursive rather than material goal—"an AI system is only truly regarded as 'working' when its operation can be *narrated in intentional vocabulary*, using words whose meanings go beyond the mathematical structures. When an AI system 'works' in this broader sense, it is clearly a discursive construction, not just a mathematical fact, and the discursive construction succeeds only if the community assents" (Agre 1997b: 14, added emphasis). He then explains the difficulties of AI researchers in addressing the conflations and disparities of signifiers with their objects—the external reality:

> "The point, in any case, is that the practical reality with which AI people struggle in their work is not just 'the world,' considered as something objective and external to the research. It is much more complicated than this, a hybrid of physical reality and discursive construction. The trajectory of AI research can be shaped by the limitations of the physical world—the speed of light, the three dimensions of space, cosmic rays that disrupt memory chips—and it can also be shaped by the limitations of the discursive world—the available stock of vocabulary, metaphors, and narrative conventions." (Agre 1997b: 15)

Linguistic-philosophical perspectives often seem to be operating under an implicit assumption that our discourse is limited to discussing entities that exist in the real world, things that we believe to exist, or other kinds of valid descriptions of reality. When it comes to the scientific discourse, this assumption may indeed hold true—at least, this is the ideal situation.

However, when using language in everyday situations, people first and foremost simply speak to express themselves, and the relationship between what is said and the extratextual reality takes a secondary place. With the rise of phatic technologies such as social media applications and other systems enabling the sharing of nonverbal reactions (emoji), memes etc, it has even been suggested that the phatic function of language takes precedence over other functions—people simply interact to keep in touch.

As discussed in Paper III, the logic-philosophical status of a statement regarding imaginary objects may indeed be that it is about nothing, or nonsensical, but such knowledge can never be derived from the text itself. In *A semiotic-pragmatic approach to literature*, Jørgen Dines Johansen points out that "On a basic level, the linguistic representation, at least in English, is identical regardless of whether the universe referred to is fictional or historical (e.g., verb tense is no certain indicator)" (Johansen 2002: 152). Marie-Laure Ryan criticizes the literary theories relying on Fregean logic, according to which "reference can only be made to that which exists" in the real world, which is ultimately the only world (Ryan 1991: 14).

In semiotics, Thomas Sebeok has considered a similar perspective:

> "Those of us who practice semiotics tend to treat these happenings the same way despite their manifest substantive differences of setting, cast of human or speechless characters, and many other variables. What entitles us to do so is an abstractive operation which resolves each episode to an instance of semiosis, or sign action. In this view, semiotics is not about the "real" world at all, but about complementary

> or alternative actual models of it and—as Leibniz thought—about an infinite number of anthropologically conceivable possible worlds. Thus semiotics never reveals what the world is, but circumscribes what we can know about it; in other words, what a semiotic model depicts is not "reality" as such, but nature as unveiled by our method of questioning." (Sebeok 1991: 12, 2001: 26)

Similarly, Umberto Eco indicates that our experience of the reality is foremost semiotic: "Within the framework of a constructivistic approach to possible worlds, even a so-called 'actual' or 'real' world of reference must be taken as a possible world, that is, as a cultural construct" (Eco 1979: 222).

In this context, the underlying theories often cited in computational linguistics and computer science—such as Frege's predicate logic etc, according to which statements about non-existent objects simply *do not refer* (Paper III)—seem quite naïve. This naivety is reflected in the discursive surprise on how easily LLMs seem to 'hallucinate', or produce statements of things, phenomena, and resources that do not exist in the real world. Such a view reflects the deep misunderstanding of the functions of language that seems to be widespread among computer scientists. It also misses the point that LLMs are also trained on texts of literary fiction which use the same level of grammatical confidence. Eventually, language is a symbol-based semiotic activity, and "the symbolic level of meaning-making, predominant for cultural systems, tends to close itself off from the extra-semiotic reality" (Rickberg 2023: 13).

The problem is not about finding the proper place where meanings become unanchored from their signifiers and divorced from the reality during the synthetic text generation process. The fallacy starts with the very expectation that there ought to be such a connection at all. There has never been any relation between text and reality in the first place—at least not any kind that is not embodied, lived, and deeply embedded in human beings and our everyday social experience. Indeed, meanings and their (temporary) objectivations—that can be relocated between human beings—are simply by-products in the communicative process that is *semiosis*. And the ways in which language models fail to meet their users' expectations simply reveal that, contrary to our habitual beliefs, perhaps language is not the best device for conveying meaning and information at all.

"All fiction is metaphor. Science fiction is metaphor. What sets it apart from older forms of fiction seems to be its use of new metaphors, drawn from certain great dominants of our contemporary life— science, all the sciences, and technology, and the relativistic and the historical outlook, among them. Space travel is one of these metaphors; so is an alternative society, an alternative biology; the future is another. The future, in fiction, is a metaphor."

— Ursula K. Le Guin,
The Left Hand of Darkness

2. THE EXPECTED AI AND ITS SOCIOCULTURAL FUNCTIONING

Researchers increasingly agree that the exact meaning of AI is "a matter of some debate and continual redefinition" (Gray 1997: 128). Throughout this work and associated papers, I do not explicitly define AI but allude to the idea via its cultural representations and manifestations. These objectivations include GPT-3 (Generative Pre-trained Transformer[6] 3), the large language model (LLM) built by OpenAI, the derivatives of which underlie the currently most popular and publicly available "chatbot" ChatGPT. The semiotic perspectives on the reception of GPT-3 outputs in culture are reflected in Paper II and Paper III. Another analysed objectivation of AI includes the fictional robot, which I address in Paper I. In this thesis, I also use the concept of AI in a manner that reflects its usage in public discourse—roughly as an umbrella term for a wide variety of technologies, ideas, and representations. Eventually, I argue—as shown above and below—that AI has no fixed meaning at all. Instead, AI in discourse means whatever the respective author or reader wants it to mean; and often, when the term is used rather as a sensationalist keyword, it has no real meaning at all[7].

Following its discursive use and overdetermination with various meanings and referents, the concept of AI is a prime example of an empty or floating signifier.

> "In the Essex tradition, discourse is the primary terrain of objectivity as such and the problem of the constitution of social reality becomes the problem of the constitution of discourse. We can broadly describe the process as follows. Various signifiers float in a discursive field; their specific meaning is not fixed. They are, so to speak, overflown, or overdetermined with meaning. At some point, a signifier acquires a dominant position in this discursive field and establishes the identity of the floating signifiers by fixing them in a meaningful whole, that is, the heterogeneous elements of the discursive field become moments of a specific discourse." (Selg and Ventsel 2022: 672–73) (see also Laclau 2005)

The idea of an empty signifier first appeared in the works of Lévi-Strauss (1950: 110) as a 'floating signifier' (*signifier flottant*), defined as "the disability of all finite

---

[6] Transformer architecture was first introduced by Vaswani et al. (2017).

[7] To demonstrate this, Daniel Leufer and colleagues have built a website aimyths.org where one can play with a "terminology replacement" tool. The tool enables experimentation by replacing the phrase "artificial intelligence" with another phrase of choice, such as "complex technology" etc. The authors argue that if, following the change, the sentence does not make any sense, it is a good indication that the term AI in this context is meaningless as well. The terminology replacement game is only one tool helping to decrease the anthropomorphism regarding AI.

thought”[8] (Lévi-Strauss 1987: 63). In deconstructionism, empty signifier refers to a situation where the signified is very vague, difficult to determine or non-existent: “Such signifiers mean different things to different people: they may signify many or even any signifieds; they may mean whatever their interpreters want them to mean” (Chandler 2017: 90).

According to my Tartu semiotics colleagues, “the ‘empty signifiers’ that start to dominate certain discourses as their centres, do not come out of ‘thin air’ but have their own semantic, pragmatic, and historical identity that starts to become more and more ‘empty’ as they become the centers of particular discourses” (Selg and Ventsel 2022: 4). Revealing this historical composition and identity of AI is one of the goals of this thesis.

Artificial intelligence is also an empty signifier due to being a symbol-type sign (in the Peircean sense). As such, it is not meant to have a purely physical grounding (such as referring to a specific tool or program). Instead, the concept of AI refers to “a cluster of other sign relations” (Favareau 2015; Rickberg 2023: 9). As biosemiotician Donald Favareau argues, “the ‘ground’ of symbolic reference in a sense *depends* on such symbols never unilaterally resolving into a single, fixed, intellectual entity or concept” (Favareau 2015: 253). This is also why and how “the symbolic level of meaning-making, predominant for cultural systems, tends to close itself off from the extra-semiotic reality” (Rickberg 2023: 9).

In this sense, the empty or floating signifier possesses the highest level of symbolicity, being far removed from its referents and disconnected from the extra-semiotic reality. It is in this manner that the signifier *artificial intelligence* can operate at once as a reference to almost any computational object or a cluster of ideas, and as an *AI effect* acting upon the discourse itself. The latter concept comes from a quote attributed to computer scientist Lawrence Gordon Tesler, colloquially repeated as “AI is whatever hasn’t been done yet.”[9] This attitude characterizes a large part of the developments in computer science, echoed by engineers and AI researchers alike. Therefore, it seems that the very definition of AI has, since the very beginning, contained a “miraculous component” that Ashby cautions against in his discussion of the Black Box concept.[10] Thus, any ‘black

---

[8] « Nous croyons que les notions de type *mana*, aussi diverses qu’elles puissent être, et en les envisageant dans leur fonction la plus générale (qui, nous l’avons vu, ne disparaît pas dans notre mentalité et dans notre forme de société) représentent précisément ce *signifiant flottant*, qui est la servitude de toute pensée finie (mais aussi le gage de tout art, toute poésie, toute invention mythique et esthétique), bien que la connaissance scientifique soit capable, sinon de l’étancher, au moins de le discipliner partiellement. La pensée magique offre d’ailleurs d’autres méthodes de canalisation, avec d’autres résultats, et ces méthodes peuvent fort bien coexister. » (Lévi-Strauss 1950: 110, original emphases)

[9] Tesler himself argued that he was misquoted and actually had said that “Intelligence is whatever machines haven’t done yet.” *Larry Tesler Consulting website, CV: Adages & Coinages*, http://www.nomodes.com/Larry_Tesler_Consulting/Adages_and_Coinages.html. Accessed September 8, 2023.

[10] Ashby’s treatment of the miraculous and the Black Box are discussed more in depth in Chapter 2.1.1.

box' system seems like something extraordinary, but as soon as researchers figure out the real configuration of parameters within, they lose interest in calling it AI and label it with a more specific term. The concept of AI at the same time retains the connotation of being a continuously elusive future possibility.

As soon as the definition of AI starts to solidify, the discourse takes care to remove any concrete referent objects from its reach again.

* * *

Many studies have revealed the extent to which technologies, especially AI, borrow their meanings and reference objects from science fiction (Dinello 2005; Cave et al. 2018, 2020; Cave and Dihal 2019), myths (Mayor 2018; Musa 2020), antique thought (Cave and Dihal 2018) and other forms of fiction.

> "It is surprising that AI is able to influence today's world to such an extraordinary extent. One possible explanation for this is an (of course not openly conceded) projection of the conception of human intelligence into machines which could be prevented by a generally accepted, neutral definition. We might wonder why human intelligence is so quickly assigned to computers. The role of AI in our life depends on our expectations in regard to the performance of these machines, and our expectations are predominantly based on our misguided conceptions of the potentiality of the machine and not on its technical performance." (Born and Born-Lechleitner 1987: viii)

Unfortunately, such studies usually add to the terminological confusion and status of AI as an empty signifier, not helping to clarify the concept further. Therefore, researchers increasingly suggest using a more precise term instead of "AI".

Save for the famous first-in-line science fiction characters such as HAL9000, C3PO etc, or Ava from Ex Machina, what, then, are we talking about when we say, "artificial intelligence"? The following subchapters describe what I perceive to be the most common aspects in the contemporary and historical cultural construction of AI as a concept and reference object.

Following the papers and an additional analysis of the connotations of AI, I propose a new framework for evaluating the possible meanings, connotata, denotata, and reference objects of AI. Current media discourse on AI contributes to the formation of a cross-cultural composite image—an "Expected AI"—as visualized on Figure 2. The Expected AI is composed of several factors. An AI system or device may be described and defined roughly as a socio-material device, a set of functions, or an idea grounded in human–machine comparison. These three parts have deep historical roots, reflecting and reiterating cultural texts and ideas that go back at least to Ancient Greece. There is also a fourth factor that is more recent and gaining increasing foothold—data. Interestingly, interpreting AI as data seems to offer a strong alternative to interpreting it as human's Other. Therefore, I suggest that the concept of *data* may overtake the human-machine comparison as these concepts seem mutually exclusive. Data is always *about* something, so when the system is about collecting and processing data, it is not so easy to picture it as personified, compared to the human being.

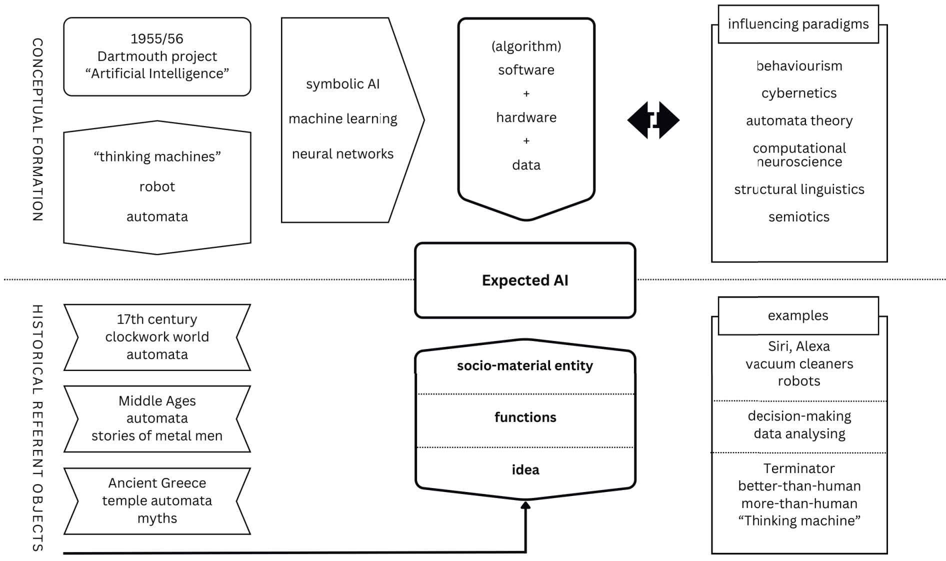

Figure 2. The historically and culturally informed construction of the Expected AI.

As the various facets of the Expected AI are crystallized and legitimized in discourses by further repetition and retelling, the imaginary Expected AI becomes a kind of constraint or attractor to any new text utilizing the term AI. The situation can also be described in terms of the Expected AI becoming a habit: “Peirce’s notion of habit entails the tendency of any process to have a tendency, to realize only some of its initial potential trajectories” (Marais 2019: 56). Semiosis as the process of meaning-making works with such trajectories which, in suitable conditions, drive new interpretations down the habitual paths. The interpretations and associated meanings of the concept can thus become the function of the discourse—the kind of “AI effect” mentioned above. Eventually, it is not even possible to draw a cause-effect relationship between the objects and their meanings. In the words of Kobus Marais, “Causes might thus sometimes be effects, and agent and structure might influence one another mutually. Trajectories could be both cause and effect, as could constraints” (Marais 2019: 62).

2.1 The history of the concept of AI and its referents

The concept of *artificial intelligence* (AI) is relatively new, dating back to the 1955 proposal of the Dartmouth project on AI (McCarthy et al. 1955, 2006). The initiators of the famous field-founding event define AI as a research field attempting to “make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves” (McCarthy et al. 2006: 12). More generally, AI was defined as the objective of “making a machine behave in ways that would be called intelligent if a human were so behaving” (Nilsson 2009: 53).

Since its conception, the original authors attempted to move the field of AI away from several fields such as psychology (behaviourism) or Wiener's cybernetics, and gain distance from the 'automata theory' popular at that time (McCarthy 1988).

> "Allen Newell (1983), in a study of the historical development of this discipline, has shown how AI, after travelling through various paradigms, finally found its own route. Indeed, since its origin in the 1950s, AI has distinguished itself in turns from cybernetics, systems theory, pattern recognition, numerical computing, programming theory, electrical engineering, and finally, formal logic." (Meunier 1989: 44)

Despite such positioning, the AI paradigm has since relied on and contributed above all to the computational models of society, culture, and mind. The Dartmouth project was also the first to insist on focusing AI on software (programming) rather than hardware (machines, robots) (McCarthy 1988: 227).

Historian of technology Derek de Solla Price (1964) suggests that complex technologies result from preceding mechanistic philosophy, rather than the other way around. Consequently, the sociocultural inclination "toward mechanistic explanation led to the making of automata," further serving as a foundation for subsequent technologies, scientific instruments, and the Industrial Revolution. The idea of 'living automata' was additionally supported by the concept of *Natura artifex* popular in the 17th century. The idea of Nature as a forger of mechanical things and the ideas about the mechanics of life and 'clockwork world' were present in the works of Thomas Hobbes, René Descartes, and Robert Boyle (Truitt 2015).

Eventually, "In the augmenting success of automata through the age of Descartes, and perhaps up to and including the age of electronic computers, we see the prime tangible manifestation of the triumph of rational, mechanistic explanation over those of the vitalists and theologians" (de Solla Price 1964: 10). In other words, first the world was described as a giant clock, and later—even until now—attempts persist to remake the world in this image.

When talking about the introduction of the concept of AI in 1955/56, it is important to recognise that the idea did not arise out of thin air. This is why I prefer to include all possible mechanical and automated constructs throughout history as possible referent objects for AI. The previous cultural knowledge of these objects and related texts—picked up and repeated over and over again—strongly contribute to our contemporary understanding of AI. These referent objects will be discussed in the following chapter.

Today, the definitional focus of AI systems in academic discourse has shifted to data, such as in the definition offered in the 2019 special issue on AI by the Royal United Service Institute (RUSI) journal. According to this definition, "AI refers to the output of three interacting elements: computer hardware, […] software, often referred to as algorithms […], and data […] or statistics [comprising] collections of 'facts' that are seen to be sufficiently similar as to be meaningfully aggregable" (Taylor 2019: 73). Further on, the author gives three categories for possible AI outputs: probabilistic forecasts; prescriptions, or recommendations for action; and the capacity to act or implement the prescription (Taylor 2019: 74). This division also draws a clear distinction between AI and autonomy: AI systems can be used to diagnose, forecast, or give recommendations for action, without the system having any action autonomy at all.

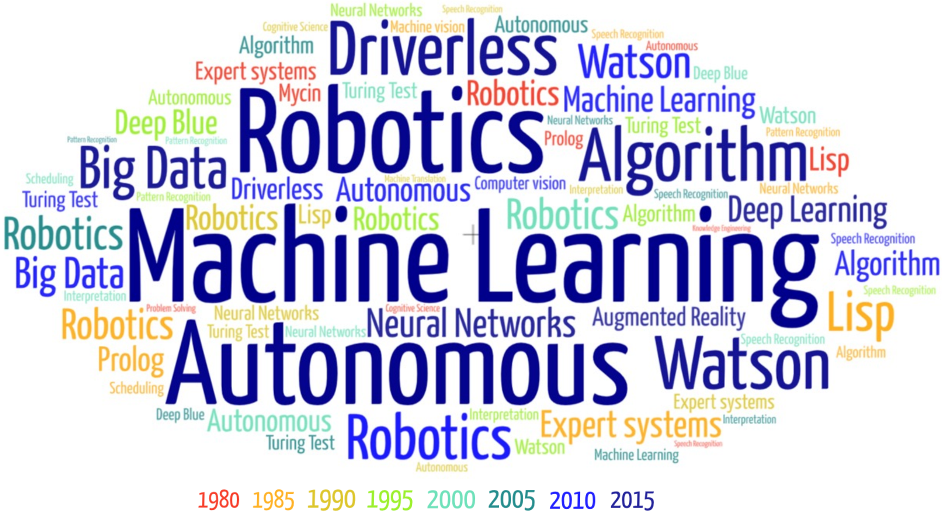

Figure 3. This word cloud visualises the most popular AI-related keywords in relation to their frequency (the bigger the word, the more frequent). Generated by the author, based on the research by Zhai et al (2020: 143) who analysed articles from five major US and UK media outlets (approximate period 1980–2015), extracting keywords related to AI technologies.

One of the popular methods in recent research is the mapping of AI-related concepts through the analysis of media texts. Some rely on quantitative rather than qualitative methods (e.g., Fast and Horvitz 2017; Zhai et al. 2020; Zeng et al. 2020), thus only the most popular or frequent keywords find their way into the published paper. Bearing in mind the limitations of the studies, even the presentation of the most popular keywords over time provides an interesting overview. Zhai et al. (2020) surveyed five major anglophone media outlets (*The New York Times, CNN, The Guardian, USA Today, The Washington Post*) for select keywords related to AI. Figure 3 displays an alternative visualisation of found keywords as a word cloud. The colour of the keyword corresponds to the relative time frame (the authors divided their results in approximate 5-year periods). However, this must be placed in the context of the total number of the provided keywords. Figure 4 displays the same keywords (Zhai et al. 2020: 143) with their absolute frequencies. In absolute terms, the keywords ranked 7th to 10th in 2015 appear more frequently than the most popular keywords from the 1980s. At the same time, the analysis clearly shows that the last decade has seen an explosion in the number of AI-related articles and headlines in the media outlets surveyed.

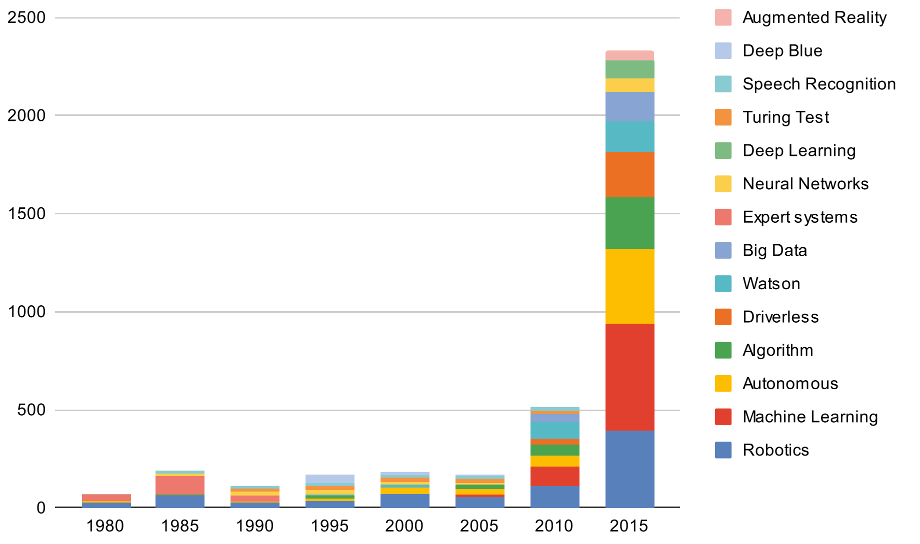

Figure 4. The endurance of AI-related keywords over time. Author's diagram, based on the given frequency of top 10 keywords provided in Zhai et al. (2020: 143).

Nevertheless, the studies provide insight into some of the more enduring connotations of AI, such as *autonomous* or *algorithm*. In line with this, current vernacular often conflates these concepts with AI, especially since the increasing usage of recommendation algorithms in various online platforms. As an example, anthropologist Beth Singler analyses tweets that mention being "blessed by the algorithm (gods)" (Singler 2020). At the same time, scientific metalanguage is not far behind (e.g., Shane 2019). Alongside driverless cars and autonomous weapon systems, Simon Chesterman brings out *algorithmic decision-making* as a distinct category (Chesterman 2021: 32). International relations professor Kenneth Payne also emphasizes that "AI is a decision-making technology, rather than a weapon" (Payne 2021: 3). Recent literature on AI and the military recommends the distinctive use of the concepts of AI and autonomy (Scharre 2018; Taylor 2019; Payne 2021).

2.1.1 θαῦμα ἰδέσθαι and the politics of the spectacle

Historians of technology like to date the "forerunners" of AI significantly earlier than 1956. McCarthy was certainly not the first to dream of making intelligent, autonomous machines—indeed, the latter idea is almost as old as history itself. The hardware problem is also still relevant to AI and grounds the concept to much older cultural history. Writers of history, when discussing robotics and AI, sometimes like to go back as far as Ancient Greece (Mayor 2018; Pickover 2019) or at least to medieval automata (Truitt 2015). Automata are often considered to form "the proto-history of robotics" (Ambrosetti 2010: 21). Indeed, Ancient Greek writings present a remarkable assortment of mechanical devices crafted to elicit awe and amazement among their audience (Tybjerg 2003: 443). The most prolific and best-known treatises are those by Heron of Alexandria, notably Ἥρωνος Ἀλεξανδρέως περὶ αὐτοματοποιητικῶν (*On Automata Making*; Latinised as *De automatis*). The manuscript details descriptions and schemata of automated constructions in religious theatre and temple mechanics. The purpose of self-moving mechanisms such as temple doors, god statues and alike, was to convince the worshipping public of the presence of deities in religious sites (Bur 2016). Further back in time, the ideas of "living statues" and other miracles appear in Ancient Greek mythology (Mayor 2018). The imaginaries of automata continue in the stories and legends of Middle Age Europe (Truitt 2004, 2015). Historically, automata were also popular in the pre-13th century Middle East–for example water clocks, water pumps and other systems described in *The Book of Knowledge of Ingenious Mechanical Devices* (Jazari 1974; Nadarajan et al. 2007). Therefore, the concept of automata has long-standing cultural roots and the connotations associated with such things have been carried over from context to context over the centuries.

At the same time, classics scholars caution against treating ancient automata as direct precursor to contemporary robotics, arguing that temple and theatre mechanics were operating in a different and very specific cultural dynamics (Bur 2016, 2020). However, various aspects and circumstances in the context of ancient automata bring effectively together technology, politics, magic, and—sometimes rather semiotic—meta-language. Most importantly, the stories about both fictional and actual automata connect the significance of the mechanical devices with the trajectory for a magical interpretation in a manner that reinforces the politics of it all.

Heron's *On Automata Making* contains engineering schemas and corresponding instructive descriptions for mainly two kinds of mechanisms, distinguishable by their function. Temple automata concern statues of gods and other stationary details such as temple doors or other parts, designed to move on their own to an extent. The second type represented in the manuscripts concerns movable automatics used in theatres (Prou 1884 provides a French translation and commentary of Heron's manuscript; Grillo 2019 provides a partial English translation). Such Ancient Greek theatre mechanics is perhaps best understood via the latinized concept *deus ex machina*, a plot move to suddenly introduce a god on stage to solve the plot problem. Not just a rhetorical device, the saying signifies the situation of the 'god' appearing quite literally from a machine (or *as* a machine). While the function of theatre may have been more recognizably fictional and entertaining, the self-moving statues, doors and other items in the temples were meant to be interpreted much more literally, as a "final indication (*sēmeion*) of the god's presence" (Bur 2016: 35). Tatiana Bur (2016) argues that the religious automata functioned to cement social and psychological structures of power via the political use of spectacle machinery; animated god statues were meant to convince people of the agency of the god-as-statue, and people—worshippers—were actively seeking out such signs that would confirm that gods have heard their prayers.

In this context, it is also important to understand that in Greek texts, the signifier—αὐτόματος—is not necessarily referring to automata as we understand it today. Instead, its first meaning—*self-moving*—takes precedence, signifying the characteristics of spontaneous events or even human behaviour, extended only to mechanical devices during the Hellenic period (Ambrosetti 2010: 5). Bur goes further, arguing that "it is more accurate to define automata as *self-animated* rather than self-moving [as these machines] acted of their own accord in terms of more than just movement, as they often included a range of sensory displays" (Bur 2016: 25, added emphasis).

Apart from the religions, Greek temple mechanics served a political purpose benefitting the rulers:

> "We cannot help but touch upon, here, the politicised aspect of these machines and the way in which they were actively harnessed as part of political theatre by leaders who wanted to bolster their own position and relation with the divine." (Bur 2016: 36)

The "political theatre" evokes a rather cynical picture of the society in question: there are ordinary people, lured to believe in the divine presence in purely human-made constructions by various forms of trickery. Heron dedicates a significant part of his descriptions accompanying the engineering designs to explaining how to achieve this impression and be left without any doubts as to its divinity:

> "The last point to note is the preoccupation with the smoothness of the display indicating, as with *On Automata Making*, that it is imperative that the viewer subscribe to the miracle so that no scepticism be aroused in the viewer." (Bur 2016: 142)

On the other side, there are political leaders attempting to manipulate people in their favour. In working for this purpose, the concept of θαῦμα ἰδέσθαι, translated as *a wonder to behold*, seems to be central. *Thauma*—the miraculous—is also a concept that directly connects the descriptions to technology: "every time the formula *thauma idesthai* is used in the Iliad it relates to *technē* and the divine" (Bur 2016: 29), confirming that "the mechanical and wondrous aspects were in fact inextricably entwined" (ibid: 39). This leads to another puzzle. Interestingly, despite the belief in the magic of technology, the audience also had a clear understanding that the devices were human made; yet this did not prevent them from functioning as vehicles for awe.

> "By transferring the unaccountable transformative power of natural or divine forces to the realm of human skill and experience, the marvelous starts to be associated with the latter. […] The technical miracle is miraculous *precisely because it is achieved by human activity*; as such it substitutes or supplements the natural." (Gerolemou 2019: 45, added emphasis)

How was such seemingly contradictory attitude possible? Tatiana Bur argues that "The automaton is an acknowledgment and a manifestation of the deity's power, in turn intimately linked to the polis', or in certain instances the polis elite's, power" (Gell 1992; Bur 2016: 44). In fact, in the context of the myth of human becoming—by that, I mean the story of Prometheus who first made humans and then bestowed various skills upon them—the extrapolation of the same principle on the situation with automata does not seem so far-fetched after all. Apparently, the technology with its miraculous complexity was taken as a sign that the gods have kindly bestowed the knowledge and power to build them upon the ruling elite who, therefore, are deserving of further support.

Strangely enough, this does not sound that different from the system of venture capital and political lobbying of the Silicon Valley.

* * *

In the Middle Ages and early Renaissance Europe, the concept of automaton disappears, being replaced with *ingenium,* pejorative *artifice,*[11] and other terms (see Truitt 2004, 2015; Ambrosetti 2010). Heron's *On Automata Making* is the most significant source for the concept of *automaton* and since its translation to Latin and re-introduction to Europe in the 16th century, the word reappears in discourse as well.

A lengthy written overview of things considered under the label *automaton* originates from a 19th century printed report *Automata Old and New* by Conrad William Cooke (1893), based on a presentation delivered in an 1891 meeting of Ye Sette of Odd Volumes, a London-based literary and esoteric dining club.

> "The word Automaton would in its strictest and most comprehensive sense include all apparently self-moving machines or devices which contain within themselves their own motive power, and in this sense such machines as clocks and watches, and even locomotives and steamships might be included. I shall, however, throughout this paper limit myself to the more restricted and more ordinarily accepted meaning of the term, namely, such self-moving machines as are made either in the forms of men or of animals, or by which animal motions and functions are more or less imitated." (Cooke 1893: 12–13)

Cooke (1893) goes on to discuss "self-moving machines" from the moving statues made by Hephaestus in Homer's Iliad and Daedalus' mechanical wonders to the more historically supported accounts of 17th–18th century mechanics. In his report, references to specific books and historical figures are mixed both with explicit

---

11 "Two Latin terms, *ars* and *ingenium*, and their cognates in Old French, encompass a range of objects, activities, and connotations, from the skillful creator, as in *Natura artifex*, to possible demonic involvement in creating moving or speaking statues. [...] In Old French *artifice* has a range of meanings, from admirable skill to sinister fraud: it meant one who worked as an artisan and a creator who copied natural forms with the intention of fooling the senses. [...] Both *engin* and *ingenium* were used until the seventeenth century to denote a manmade mechanical device, such as an automaton." (Truitt 2015: 49–50, original emphasis)

references to mythology and uncited narratives that remind more of folklore than the history of science, illustrated by the following, for instance:

> "Then again the mighty Odin had among his mystical possessions a speaking head, believed to be that of Minos, which Odin preserved by encasing it in solid gold. He is said to have consulted it on all occasions, and its utterances were regarded as oracles." (Cooke 1893: 47)

Cooke estimates that "The great period for the construction of automata began at the close of the fourteenth century and reached its climax at the end of the seventeenth and beginning of the eighteenth century" (Cooke 1893: 49). Despite declaring his preference for automata "made either in the forms of men or of animals," (Cooke 1893: 12–13) Cooke spends a fair amount of space listing and detailing clocks and other mechanisms. His accounts are most detailed when it comes to the 18th century.

In Ancient Greek storytelling, just as in Cooke's account, historical facts easily merge with mythical ones. Allusions to arguments over the existence and qualification of gods can be found in most descriptions of related myths and cults, especially from 500 BC onward when "the Homeric picture of the gods no longer satisfied intellectuals" (Bremmer 1994: 89). By the fourth century BC, Greek myths were rather considered "old wives' tales" (Bremmer 1994: 65). This critical stance, however, did not stop Ancient Greeks from tracing their ancestry to certain gods whenever such connection had socioeconomic or political benefits. The second-century scholar and traveller Pausanias—whose *Periegesis* remains the best known geographical and cultural overview of his contemporary Greece—treated godhood as a form of status that sometimes extended to the land associated with the god (Pirenne-Delforge 2010). A few centuries earlier, Cicero recounts a clash between Roman tax collectors and the people in Boeotia over exempting lands "belonging to the immortal gods" from tax—which the Romans did not want to recognise, arguing that the gods under question were not real gods (Cicero 1967: 3.49; Pirenne-Delforge 2010: 382).

In the above context, it now seems significant that Trenchard More, who in his Dartmouth report dedicates significantly more pages to W. Ross Ashby than to other participants, notes that "A point which Ashby stressed several times is the following. A simple machine *appears to be extraordinary* when viewed psychologically. When part of a mechanism is concealed from observation, the behavior of the mechanism *seems remarkable*" (More 1956a: 2, added emphasis). A similar observation seems to carry over to More's own conclusions: he "approached the subject from his symmetric model of a machine and suggested that in complicated machines, the comparator might observe only machine effects on the environment, rather than the internal operation of each machine" (More 1956b: 8). In his book *An Introduction to Cybernetics*, Ashby explains: "It should be noticed that as soon as some of a system's variables become unobservable, the 'system' represented by the remainder may develop remarkable, even *miraculous*, properties," such as conjuring, "which achieves (apparently) the miraculous, *simply because not all the significant variables are observable*" (Ashby 1956: 114, added emphasis). This statement forms the crux of the Problem of the Black Box, which Ashby discusses in detail (Ashby 1956: 86–117). Of course, for Ashby, all (real) things are black boxes, making the theory of the Black Box "the theory of real objects or systems" or "simply the study of the relations between the experimenter and his environment, when special attention is given to the flow of information" (Ashby 1956: 110).

Indeed, when translated into contemporary context and applied to the current discourse on AI most widely, the power of the miraculous is persisting. Just as observed with Heron's automata, similar beliefs seem to occur regarding contemporary robotics. Today, devices such as the robot Sophia[12]—a "talking" android built by Hanson Robotics—despite being quite robust and obviously human-made, seems to provoke similar awe as the temple miracles did in Ancient Greece.

The Greek concept of *thauma* is "identified through the disruption of an obvious relationship between cause and effect" (Gerolemou 2019: 45). A similar phenomenon seems at work when computer scientists or engineers are surprised at the (unexpected) results received from an AI device—and discursively, this surprise is often objectivated as "emergent properties".

I suggest that the discourse on the "miraculous" is, above all, a political one. In this argument, I rely on Langdon Winner's analysis of the social and political implications of technology. Winner argues for a theory of technological politics that "draws attention to the momentum of large-scale sociotechnical systems, to the response of modern societies to certain technological imperatives, and to the all too common signs of the adaptation of human ends to technical means" (Winner 1980: 123). Formulated considerably before the latest developments in AI, his points remain valid today and for the foreseeable future.

In his widely cited paper *Do artifacts have politics?,* Langdon Winner draws attention to the ways that technical arrangements constitute social order through deliberate design choices. He examines the social and political impact of design on the example of Robert Moses, the urban planner who shaped New York City architecture from 1920s to 1970s. Winner argues that Moses had Long Island overpasses deliberately engineered to prevent higher vehicles such as buses from accessing certain areas, effectively excluding from Long Island lower-income individuals who relied on public transportation. His design choices—arguably reflecting his own racial prejudice and social-class bias—have consequences well beyond his own life and sociocultural context, embodying "a systematic social inequality, a way of engineering relationships among people that, after a time, becomes just another part of the landscape" (Winner 1980: 123–24). As the study of the history of technology shows, new tools are not always oriented for efficiency or better resource management. Unfortunately, "Technological change expresses a panoply of human motives, not the least of which is the desire of some to have dominion over others, even though it may require an occasional sacrifice of cost-cutting and some violence to the norm of getting more from less" (Winner 1980: 124).

---

[12] Hanson Robotics website describes Sophia as "the world's first robot citizen." — https://www.hansonrobotics.com/sophia/. Accessed October 1, 2023.

Anthropologist Alf Hornborg criticizes Winner's approach in that the latter "is concerned with how technologies shape the lives, minds and societies of those who have access to them, rather than with how they shape *global* relations between those who do and those who don't" (Hornborg 2021: 216, original emphasis). This is a specific shortcoming that I will address in Chapter 2.4 with a proposal for initial remodelling and further research on the full-scale socio-material figuration that is AI.

What the use and reception of technology can tell us about the intentions of its creators is another matter of inquiry. Winner (1980) argues that the overpasses of Long Island reflect the segregationist racism of Robert Moses and his era. Subsequent STS scholars and papers sought to cast doubt on Winner's arguments, framing it as a prime example of technological deterministic thought. Bruno Latour reframes the entire argument line around discrimination as a problem of affordances and constraints of sorts, pointing out that all kinds of technology—such as revolving doors, for instance—discriminate against certain types of acts and people (elderly, "furniture removers and in general everyone with packages") while affording certain other actions and usages (outlined in Johnson 1988 who is actually Latour; summarized and analysed in Joerges 1999: 414). While Latour is not wrong, he is failing to see the real concerns behind Winner's argument—concerns that rather "have to do with the proliferation of artificial devices in modern life, and how they either support democracy or discourage it; whether they make possibilities for people to participate in shaping the environments around them more likely or less likely."[13] Winner tries to demonstrate that technologies have potential material and social consequences well beyond their immediate time and context. Thus, Long Island overpasses exemplify how the design choices made by a single person impact other people and communities beyond his own lifetime by objectivating, institutionalising and legitimising certain practices, attitudes, and policies which then, over time, become naturalized, inevitable, unquestionable, and taken-for-granted. Placed in this context, the bridges demonstrate how racism is systemic and structural. And the subsequent argument reveals the ways how we[14] refuse to recognize these crystallised "normalities" later in time.

The discursive and political choices made involving AI have implications similar to Winner's bridges: the choices made by a select few have considerable political and social impact beyond the immediate use of these technologies. This impact is more severe for underprivileged people, and the impact of reorganizing

---

[13] Winner explains this in a documentary dedicated to the scholarly argument around the bridges. — Shahab Mihandoust and Francesco Garutti, dirs, 2014. *Misleading Innocence (Tracing What a Bridge Can Do)*. CCA. https://www.youtube.com/watch?v=0u6zYcci_5w, 30:05 (Accessed October 1, 2023). Latour and Woolgar argue in the film that it is very difficult to infer the creators' true attitudes and intentions from the technology itself.

[14] "We" refers to white people, as in the above film the argument is represented by three white male scholars.

social institutions in a certain manner may occur and last beyond what is perceived as the immediate design process.[15]

2.1.2 Normalizing discourse: narratives and intertexts of AI

Narratives and myths play an important role when it comes to public imagining and understanding the functioning of "intelligent" technologies. Portrayals and perceptions of AI can have both positive and negative impact on the discourse and public policies regarding technological development and regulations, finds a 2018 report by the Royal Society of the UK (Cave et al. 2018). On the positive side, AI narratives can inspire the ongoing work on AI in public and private sectors, help debate alternative futures, and provide a much-needed simplified understanding of technology for non-experts via the use of metaphors and tropes. On the negative side, however, narratives can also inspire fear and exaggerated expectations of technologies and foster false perceptions that are hard to overturn. Over-emphasizing humanoid representations produces a misinformed debate that has consequences for AI research, funding, regulation, and reception (Cave et al. 2018).

Unfortunately, "Most people's conception of what a robot is appears to be largely based on the way robots are depicted in fiction [—] as independent, autonomous actors that have a 'mind of their own', with a humanoid or anthropomorphic appearance" (Harbers et al. 2017). Narratives legitimize and normalize the order of things as they "render events plausible by placing them into coherent sequences" (Ritivoi 2009: 35). Fictional narratives about humanoid robots exhibiting human behaviour normalise the idea that this is what robots ought to be like. If the real-world robots still do not comply with the prescription of fiction, people still expect certain properties from the real devices, and when these do not deliver, the expectations are not met, and people are disappointed. Thus, the Expected AI still forms and constructs our expectations to an extent, even while we are semi-aware that our expectations are fictional.

> "Narratives of intelligent machines matter because they form the backdrop against which AI systems are being developed, and against which these developments are interpreted and assessed. Those who are engaged with AI either as researchers or regulators are therefore rightfully concerned, for instance, about the fact that the dominant contemporary imaginings of AI, primarily those of Hollywood cinema and popular news coverage, are often out of kilter with the present state of the technology." (Cave et al. 2020: 7)

Indeed, although a part of the discourse on AI sounds like a deliberate, premeditated, and desirable projection of science fiction plots upon reality, not all AI researchers agree with the tendency of computer science to venture in the realm of the fictional. Most notably, Philip E. Agre spent a large part of the 1990s

[15] The increasing use of cobalt in rechargeable devices is an example that I discuss in section 2.4.

challenging the tendency of AI researchers to conflate representations with the reality to the point of representationalist fallacy. More recently, the metalinguistic critique of AI is objectivated in ironic statements such as "When you're fund-raising, it's AI. When you're hiring, it's ML. When you're implementing, it's logistic regression" or wishing "we could stop using such an empty, sensationalized term to refer to real technological techniques."[16]

As for the correspondence to the present state of technology, the Moore's Law provides a useful example. This techno-deterministic "law" claims that the computing power of machines doubles every 18 months, which is one of the cornerstone arguments for Artificial General Intelligence and "singularity" hypotheses. Commenting on this infinite exponential growth hypothesis of computing power that is associated with the correspondingly shrinking size of processors, Estonian AI researchers argue that "These laws are about to expire as the physical limits to the reduction of transistors and the interconnections between them have been reached" (Tõugu 2018: 58) and the size of chips cannot become smaller than the atom (Laan 2017). Recontextualising Moore's Law for the study of myths and narratives, sociologist Vincent Mosco has ironically rephrased it as "Gore's Law": "Myths about the Internet double in their distance from reality every 18 months" (Mosco 2004: 1).

But what exactly is the relationship between science fiction and the reality? Literary scholars and science fiction writers alike have argued that the entire genre of science fiction, by definition, ought to be about 'other worlds', that is, not about any known reality at all, and described as so different from the present that no logical associations can be made between fiction and the external reality of the reader (Tomberg 2019, 2023). This aim is contrasted to realism as "an artistic technique that seeks to convey reality as faithfully as possible" (Tomberg 2023: 23). Still, science fiction constructs the kind of otherness that is believable *within its storyworld*—which consists of imaginary contexts that are mentally construed by the reader, based on the "alternativity markers" provided by the author (Csicsery-Ronay 2004: 123; Tomberg 2023: 95). At the same time, researchers have observed repeating trends where works of science fiction are approaching literary realism: one such trend appeared in the 1970s, and another is happening right now. The previous trend prompted science fiction writer Ursula K. Le Guin to add an *Introduction* to the second edition of her famous novel *The Left Hand of Darkness* (1976). This *Introduction*, represented in various epigraphs scattered around the present thesis, tries to make it clear that science fiction is not about the future, that it merely employs future and various other constructs as metaphors to outline, criticize or reflect on issues in the contemporary present of the extratextual reality of the reader (Le Guin 2010). The present trend is characterised by the kind of poetics that is "at the same time and indistinguishably both science fictional and realist" (Tomberg 2023: 50). Literary scholar Seo-Young Chu argues that "all

[16] Joe Davison, "No, Machine Learning Is Not Just Glorified Statistics." *Medium* (blog). June 27, 2018. https://towardsdatascience.com/no-machine-learning-is-not-just-glorified-statistics-26d3952234e3. Accessed September 8, 2023.

representation is to some degree science-fictional because all reality is to some degree cognitively estranging" and that the difference between science fiction and realism is only within the degree of effort needed to represent its cognitively estranging referents (Chu 2010: 7). Hence, realism "is actually a 'weak' or low-intensity variety of science fiction, one that requires relatively little energy to accomplish its representational task insofar as its referents (e.g., softballs) are readily susceptible to representation," while science fiction "requires astronomical levels of energy to accomplish its representational task insofar as its referents (e.g., cyberspace) elaborately defy straightforward representation" and they form part of the same continuum (Chu 2010: 7). The traditional theory about science fiction relies on the presumptions that sci-fi is the genre of cultural change, and that the future is different from the present. Currently, however, the speed and intensity of technological changes is outpacing our perceptual capacity to adapt to these changes (Tomberg 2023: 36). Additionally, we have not been able to spend the necessary time, working out interpretations and cultural understandings of these somewhat novel objects and processes. In Lotmanian (2009) terms, we are living in a constant stage of cultural explosion when it comes to technological change, and we often lack the necessary time to gain perceptual distance needed to make sense of it all.

Occasionally, new technological developments—at least on a very superficial level—also appear to resemble certain mythical and past sci-fi narratives. Contemporary sci-fi literature, in turn, employs more and more artistic devices from the realist genre. Jaak Tomberg dedicates a large part of his analysis to William Gibson's Blue Ant Trilogy of the 2000s and how it operates with sci-fi realism, conflating the novels' spacetime with the immediate present and utilizing various contemporary intertexts such as references to well-known brands (Tomberg 2023). I observe a similar tendency in the narratives of sci-fi TV series featuring robot characters (Paper I). The analysed series, broadcast in the past decade, differ from previous "robot fiction" by placing storylines in contemporary realist settings as well. No more are the robots presented in far future cosmic space colonisation, but they are among us, here and now, functioning as entertainment or household devices. Unlike R2D2 and C3PO from Star Wars, contemporary robot characters are fully represented and played by human actors. Especially the choice of actors—with their full bodies, not just the voices—is particularly conducive to comparing robot characters to human characters; they are all *played by people*. Consequently, the necessary cognitive effort for imagining the "estranging referents" becomes smaller and smaller. Thus, already at the level of poetics, favourable conditions appear for confusing science fiction narratives with descriptions of reality.

However, the way we perceive certain fictional objects—their film or stage embodiments in any given period—does not fundamentally change the ontology of literature and its function in society. Either way, science fiction is not about the future, and it is not about the future of technology. And when our favourite cultural representations make it really difficult to remember this distinction, it needs to be repeated until we understand it again.

Once the science-fictional AI is placed in the context of the Enhanced Human or on the Superhuman continuum, while respecting the function of fiction in culture as a critique of social issues, it becomes possible to see the fictional robot only as a reflection on the human Other. As a representation of human properties and cultural dreams of solving various vulnerabilities of the human body, the claimed material composition of the fictional robot in any given storyworld becomes less significant than its operational and aspirational behaviour. Placed on the continuum with other superhuman, magical and cyborg characters, the fictional AI depicts very human dreams of decreased vulnerability, immortality, and the ability to extend the boundaries of physical and psychological realms. As I explain in Paper I, the fictional superhuman carries only a very small degree of difference from a regular action hero whose flexible and resilient body is detached from its everyday functions and "reduced to an insignificant shell for the mind as the 'centre of operations'," which, when extended to robots, "can be endlessly repaired or replaced" (Viidalepp 2020: 30) (Paper I).

But more often than not, the fictional superhuman, robot, or cyborg fills the role of a *magical helper* to a human protagonist, thus fitting into a story function that is significantly older than modern technohistory.

For ordinary people, the everyday experience and functionality of AI typically entails assistance in daily tasks, offering personalized experience, or providing recommendations for further action. Virtual assistants such as Siri or Alexa assist with scheduling tasks and search for information; recommendation algorithms employed by Netflix, Amazon and Spotify offer directions in further content consumption. These assistive functions of AI align with the functions ascribed to magic(al) helpers in folklore narratives, as described by folklorist Vladimir Propp (1984, 2012: 162–67). "One of the most important attributes of a helper is his prophetic wisdom: the prophetic horse, the prophetic wife, the wise lad, etc" (Propp 1968: 83). This corresponds well to the expectations people have toward their 'smart' pocket devices, as well as to how AI and robot characters are depicted in fiction (Paper I). In his *Morphology of a Folktale*, Propp examines the meaning of folktales by analysing the functions performed by their structural elements. Similar functions are found and recognisable also in contemporary tales, allowing for comparison between the old and new stories.

> "With the magical helper or a helping device, the magical tale arrives at its peak. From that moment, the end is in sight – the hero moves towards the goal and knows that he will achieve it. /…/ From now on, the hero has an entirely passive role – the helper does everything for him."[17] (Propp 2022: 220)

[17] „Kangelasele võluvahendi andmisega saavutab muinasjutt haripunkti. Sellest hetkest peale on lõpp juba aimatav. [...] Kangelane liigub nüüd kindlalt oma eesmärgi suunas ja teab, et saavutab selle. [...] Edaspidi etendab kangelane täiesti passiivset rolli. Kõike teeb tema eest abiline või siis võtab ta appi võluvahendi. [...] Abiline väljendab tema jõudu ja võimeid" (Propp 2022: 220).

Propp considers magical helpers and assisting devices as manifestations of the same function or set of functions. The helpers enhance the hero's cognitive abilities, speed of movement, give decision-making support, and enable them to transcend the boundaries between realms, including visiting the land of the dead and returning. In popular imagination, the Expected AI operates in a comparable manner. Numerous science fiction narratives, including those analysed in Paper I, depict robots that closely align with Propp's magical helper archetype in its prophetic abilities. These robots assist humans by enhancing their intellectual capabilities, aiding in decision-making, and performing various tasks such as information analysis and generating optimal action sequences.[18] Consequently, the human character becomes a passive recipient of the robot's instructions, of which the human is solely required to follow.

When we employ everyday technological tools in our lives, such as smartphone maps for navigation, internet search engines for information retrieval, or automated alerts for receiving the latest news or academic publications of interest, we can observe a similar underlying magical helper sentiment at play. It does not help the AI discourse at all that Propp's magical agent can play a dual role in the same narrative: "One and the same person can play one role in the first move and quite another role in the second (a devil as helper in the first move, but as villain in the second, etc.)" (Propp 1968: 86).

The concept of "thinking machines" was not uncommon prior to Dartmouth conference, being well spread in the popular culture and science fiction. The academia was introduced to the idea by Turing's seminal paper (Turing 1950) which raised several questions about estimating machine "intelligence" that the field is still grappling with today. Turing's famous question *Can machines think?* has been viewed as ironic, provocative and something that should not be taken literally[19]. Leonard Pinsky's short and sarcastic (but entertaining) response to Turing's paper suggests that machines should be considered as thinking only after they achieve nervous breakdown while contemplating Turing's paper about thinking machines and fail to respond to psychotherapy (Pinsky 1951). Gilbert Ryle commissioned a serious response from British philosopher Wolfe Mays (1952), but then refused to print it, presumably because Mays' line of argumentation was resurrecting all the animistic ghosts that Ryle had just attempted to exorcise in *The Concept of Mind* (Ryle 1949) under the thesis of "what he called the category mistake or reification" (Mays 2001: 4). Turing, however, brought out more counterarguments than he did supporting arguments to his views on mind and thought, so Mays claims Turing's position as speculative from the start. Apparently, Michael Polanyi was concerned with Turing's "thinking machines" and used his polemics with Turing (documented in the seminar "Mind and the

[18] For example, the prediction and action sequence of Arisa, or Dolores, described in Paper I, to name a few.

[19] Even Turing himself declared the question "too meaningless to deserve discussion" (Turing 1950: 442).

Computing Machine", 1949, Manchester) to develop his later theories on personal and tacit knowledge (Polanyi 1958, 1966; Mays 2001: 5, 21).

One of the most typical literary metaphors for the ambitions of AI is the golem. Apparently, the story of golem and Rabbi Loew is linked not only to the concept of AI, but also to the scientists themselves:

> "Curiously enough, several present-day researchers in artificial intelligence have told me that they grew up with a family tradition that they are descendants of Rabbi Loew, though they doubt this belief has had much influence. Among them are Marvin Minsky and Joel Moses of M.I.T. Further, Moses tells me that a number of other American scientists have considered themselves to be descendants of Rabbi Loew, including John von Neumann, the computer pioneer, and Norbert Wiener, who coined the term cybernetics […]." (McCorduck 2004: 15–16; also cited in Musa 2020: 1011)

The story of the "Great Rabbi Loew", incidentally, is literary fiction from the 20th century (Scholem 1969: 189, 1978: 354). The golem as the magical helper only emerges in the writings of the 16th century German Hasidim. Only after that, combined with Paracelsus' homunculus and making its way through the annals of history, the golem seems to morph into the contemporary character of the demonic servant who becomes dangerous and "destroys the world, or in any case does a good deal of damage" (Scholem 1969: 197–202).

In Tartu semiotics, composite ideas such as the golem can also be described by the concept of code-text (or text-code) in Tartu semiotics. For Lotman, "a code text is a textual system that originates from the collective memory of a particular culture, […] a textual system with a rigid syntactic order" (Madisson and Ventsel 2022: 454). "Text-codes are present in different artistic works as repeatable narrative structures; they can be seen as the most obvious manifestations of myth in the culture, or as a connecting link between myths and texts manifested in different cultural languages" (Menise 2019: 530). When it comes to the intertexts and cultural references around AI, the story of golem constitutes one such code-text. Although code-texts have an internally ordered and, to a certain extent, stable structure, the example of the golem illustrates that code-texts can and do change over time, as they gain new elements and lose other ones.

The golem has a long and fascinating history, occurring throughout the first centuries of Kabbalist texts as a creature, creation, metaphor, and a practice of golem-making (Scholem 1969: 158–204). Through textual comparison, the golem can be associated with the creation of Adam, even though the word does not appear in the Bible in this specific context. In his extensive overview of the history of golem-stories, Gershom Scholem cites a Talmudic passage (Sanhedrin 38b) describing the first twelve hours of the creation of Adam, who is called "golem"—"a still unformed mass"—before God gives him the soul (1969: 161). Additionally, "the old Jewish tradition contains several references to a tellurian earth-spirit, dwelling in Adam" (Scholem 1969: 163). This aspect serves later as a further connection between Adam, the golem and earth magic.

Another relevant story talks about the first woman that God created from the earth as well, separately from and as an equal to Adam.

> "This was Lilith, who irritated the Lord of Creation by demanding equal rights. She argued: We [Adam and I] are equal, because we both come from the earth. Whereupon they quarreled, and Lilith, bitterly disgruntled, uttered the name of God and fled to embark on her demonic career. In the third century this story seems to have been known in a somewhat different form, without the demonic Lilith. This version speaks of a 'first Eve', created independently of Adam and hence no relation of Cain and Abel, who quarreled for possession of her, whereupon God turned her back into dust." (Scholem 1969: 163)

This comparison of the created human(s) and the golem, made from the dust or the earth, becomes essential and significant in later Jewish mysticism where the act of creation is interpreted as a contract between God and Earth, which in turn can be repeated by humans by magical means. Following the Talmudic manuscripts, the creation of Adam morphs into stories of rabbis creating various things such as calves or artificial men and returning them to dust (Scholem 1969: 166). The most important source text for the subsequent magical interpretations is Sefer Yetsirah (the Book of Creation) that outlines the creation of the world (cosmos) out of twenty-two letters of the Hebrew alphabet. Scholem argues that "Though the treatise is presented as a theoretical guide to the structure of creation, it may quite conceivably have been intended also as a manual of magical practices," and that "The affinity between the linguistic theory set forth in the book and the fundamental magical belief in the power of letters and words is obvious" (1969: 169). What exactly constitutes a golem differs from interpretation to interpretation. But at least from the 11th century forward, the stories of the mystical circles of the Ashkenazi Hasim present an approximate code-text whereby rabbis engage in the Kabbalistic rite of golem-making by means of the letters of the alphabet that have magical power of Creation (Scholem 1969, 1978: 353). It is repeatedly stressed that such activities must not have any practical benefit: "This creation of a golem is an end in itself, a ritual of initiation into the secret of creation" (Scholem 1969: 177). If at some point the initiate does reap the rewards (for example, eats the meat), he has failed. When this occurs, the characters in the story sometimes forget everything they have learned so far. Most interestingly, in certain Kabbalist texts, "The danger is not that the golem, become autonomous, will develop overwhelming powers; it lies in the tension which the creative process arouses in the creator himself. Mistakes in carrying out the directions do not impair the golem; they destroy its creator" (Scholem 1969: 191).

Musa (2020) names *The Sorcerer's Apprentice* as one of the code-texts in the discourse on AI research. Many such code-texts are widely known and analysed at Western culture, especially in folklore studies. Not just confined to Goethe's poem, this storyline is so widespread that it has been assigned an individual number in the Aarne-Thompson-Uther (ATU) morphology of folktales. First named as "The Magician and his Pupil" (Thompson 1961: 113), ATU 325 story type can be connected to many Indo-European cultures and traditions utilizing folktales of magic. Its specific subtype, 325* is described as follows:

> "*Apprentice and Ghost*. Sorcerer's apprentice having read verse from forbidden book evokes ghost, but cannot make him disappear. When sorcerer reads verse backwards, ghost disappears." (Thompson 1961: 114, original emphasis)

This sounds almost like the dedicated description of golem-making from the Kabbalist tradition. In the most recent upgrade of the ATU index, the folktale type 325* is renamed as "The Sorcerer's Apprentice", presumably to keep up with the changing culture and tradition (Uther 2011: 209). The Proppian folktale functions are also derived from the original categorisation proposed by the Finnish folklorist Antti Aarne. Thereby, ATU 325 belongs to the wider category of supernatural adversaries (ATU 300–399) which also occur in Propp (1968). Propp's function of the magical or supernatural helper is covered by ATU types 500–599. The parent category "Tales of magic" (ATU 300–749) includes several other subtopics concerning magic and the supernatural (Uther 2011). All these subtopics contain content and references to code-texts that are relevant to the kinds of stories we tell about AI today.

AI researchers also cite various works of science fiction as direct inspiration (Musa 2020: 202), many of which are explicitly or implicitly inspired by Indo-European folktale types. Thus, the discourse on AI is yet another reminder of how we cannot escape our cultural history. Through our choices and expressions, culture reproduces itself, whether we intended it or not.

The code-texts for AI discourse have also been analysed as myths. For Vincent Mosco, myths elevate people beyond "the banality of everyday life" and provide a gateway to an alternative, extraordinary reality "once characterized by the promise of the sublime" (Mosco 2004: 3). The concept of cyberspace has, since Gibson's *Neuromancer* and fuelled by phatic technologies of the 2000s, offered a direction for the digital sublime. AI as the focused, objectivated and materialised invariant enabling human extension throughout the otherwise rather elusive 'cyberspace', offers another path to the sublime. One of the most persistent of such myths in the AI discourse—both public and academic alike—is what Robert M. Geraci calls the *Apocalyptic AI*.

Geraci argues that "Apocalyptic AI advocates unite Moore's Law, which describes the rate of technical progress in computer processing speeds, to biological evolution […] as a means of assuring that the movement's predictions will come to fruition" (Geraci 2008: 140). Indeed, one of the most persistent supporting arguments used by the proponents of transhumanism and especially the Singularity[20] myth is that Moore's Law—or the increase in computing power—somehow

[20] Originating from the quasi-fictional stories by science fiction author Vernor Vinge, "Singularity" refers to the envisioned moment of "superhuman intelligence" explosion that spells the end of mankind as we know it. Economist Wim Naudé (2021) provides a critical review of the concept and related issues.

guarantees the kind of development that characterizes either type of apocalypticism (transhumanism as the utopian and singularity as the dystopian imaginary of the future)[21].

Beth Singler considers the apocalypticism surrounding technology discourse as "AI anxiety" (Singler 2019). Relying on the work of Déborah Danowski and Eduardo Viveiros de Castro, Tartu semioticians Ott Puumeister and Silver Rattasepp (2022) characterise climate- and technology-related apocalyptic ideologies as a form of accelerationism:

> "The worker qua cognitive machine plugged into the Web, zombified by the continuous administration of chemical and semiotic drugs, the permanently indebted "prosumer" of the Immaterial, avidly enjoying its own exploitation, is the new heroic anti-subject of this jubilant dystopia of a frenetically devitalized post-world." (Danowski and Viveiros de Castro 2017: 52)

Various narratives perpetuated by Apocalyptic AI acolytes eventually find context and solace in one of the many ideologies of the TESCREAL bundle (Gebru and Torres (forthcoming)). TESCREAL is an acronym bringing together a variety of more or less (self-)organised ideologies: transhumanism, extropianism, singularitarianism, cosmism, rationalism, effective altruism, and longtermism. All these are characterised by high levels of eschatology and discriminatory views, they all arise from second-wave eugenics and are quite influential in the political and financial scene. The TESCREAL ideologies also share a joint dream of AGI or digitally uploaded humans (ibid.). Some of these apocalyptic ideologies are also described by Musa (2019, 2020).

Earlier, historian of technology David Noble argued for a similar characterisation of the US technological elite:

> "These scientific and technological elites, says Noble, are overwhelmingly male and aggressively masculine in their culture and values. For that matter, in the spirit of Victor Frankenstein, the ultimate dream of many of Noble's visionaries is the creation through genetic engineering of a womanless world—the culmination of centuries of mistreatment of women in general and of female engineers and scientists in particular." (Noble 1997; Segal 2012: 188)

This line of thinking highlights an underlying problem inherent in many technologies, not just AI. This is the extensively studied problem that often technologies—allegedly changing the world for better and offering new opportunities—are in fact reproducing and reinforcing existing structures of control, power, and inequality (Benjamin 2013, 2019a, 2019b; Ekbia and Nardi 2016; Eubanks 2017; Hicks 2017; O'Neil 2017; Birhane 2021).

Further, the discourse on and around AI perpetuates what Yarden Katz calls *epistemic forgeries*—"the fictions about knowledge and human thought that help

[21] As discussed above, Moore's Law is problematic in relation to predicting the future of technology.

AI function as a technology of power" and "work in tandem to produce fear and uncertainty about an impending social transformation ushered in by machine intelligence" (Katz 2020: 94, 95). The first of these depicts AI systems as presenting a universal and objective "view from nowhere". This stance toward the universal and authoritative "problem-solving" capacity of AI has been present since the beginnings of the field. At the same time, the actual perspective can rather be characterised as "a view from *somewhere*", namely the view "of a few, technically educated, young, male, probably middleclass, probably white" (Adam 1998: 94; Katz 2020: 100). The second epistemic forgery insists that AI systems are surpassing human capabilities (Katz 2020: 103–17). This is a repeating narrative that resurfaces with the sales pitch of every new AI tool. The subsequent product testing conducted by the public then reveals quite quickly the various ways in which the product fails or is outright useless. An example of this is Meta's Galactica—a large language model trained on a corpus of scientific texts. In November 2022, Galactica was released with a claim that it would henceforth generate new scientific articles. In the hands of the public, Galactica did what language models usually do—that is, generate authoritative text with no reference to the reality—and was shut down after three days.[22]

The third epistemic forgery highlighted by Katz is more implicit in discourse and entails the idea that "truth" (or accurate descriptions of reality) can be achieved "by a computational process that operates independently of people, and whose inner workings cannot be understood by them" (Katz 2020: 117–18). The real danger of this idea lies in the propagation of the belief that this "process, once set in motion, is beyond anyone's control, yet its outcomes are superior to what any human collective can achieve through deliberation" (Katz 2020: 118). The dangers of the belief in the superiority of machine decisions will be further discussed in Chapter 2.2.2.

But foremost, this forgery serves to realign power structures and reassign accountabilities:

> "This forgery leaves only two roles for people: first, the experts have to set the conditions for said computational process to unfold, and second, everyone else must do their best to adapt to the results, while acknowledging that the procedure's logic will remain forever indecipherable. This forgery not only elevates experts—as they will wield this fantastical force—but also serves to absolve society's most powerful from responsibility for social arrangements." (Katz 2020: 118)

Looking at the concentration of technology companies in Silicon Valley and their global reach, there is almost no question of who, in that world, claims the role of

[22] Galactica was probably one of the first public examples of how language models "made up fake papers (sometimes attributing them to real authors)". See Will Douglas Heaven, "Why Meta's Latest Large Language Model Survived Only Three Days Online." *MIT Technology Review*, November 18, 2022. https://www.technologyreview.com/2022/11/18/1063487/meta-large-language-model-ai-only-survived-three-days-gpt-3-science/. Accessed October 2, 2023. Today, LLMs generating fake information is most often discussed as "hallucinating".

the experts, leaving the necessity of adaptation to the rest. This is why the question of who can speak about and for the new technologies is, indeed, a crucial one for both power and politics. (I will discuss this in more detail in Chapter 2.3.)

With such a power at stake, the powerful will likely have an interest in securing it. Accordingly, Mustafa Ali argues that “Trans-/Posthumanism can—and from a critical theoretical and/or decolonial perspective should—be viewed as a response to the phenomenon of White Crisis, one that is techno-scientific and occurs in parallel with, albeit somewhat obscured by, the more overt phenomenon of conservative ‘White Backlash’” (Ali 2019: 11).

2.2 The problems of metalanguage and conflated ontologies

Several philosophers of technology and computer scientists have recently pointed out that overpromising and systematically misleading the public about the technologies’ actual capabilities may lead to another AI ‘winter’ (Floridi 2020), “in which disappointed investors and funding agencies withdraw their support from the field” (Blackwell 2021: 379) when applications continuously fail to provide the kinds of results that are advertised by their optimistic creators. Other accounts suggest that “AI winter” may designate only the visible decrease in media hype, while the developers and companies simply continue their work under different labels (Katz 2020: 60). Nevertheless, misrepresentations of ‘smart’ and ‘intelligent’ technologies can be detrimental to the development and implementation of any kind of technology. It also diverts public discussion from necessary topics.

In the case of AI, the uncritical borrowing of concepts between different disciplines (especially between life sciences and computer science) has led to the situation of ontological confusion between the “mechanisms” of life and the workings of machines. This confusion has been highlighted as contributing to the construction of AI myth in the 1950–1970s US media as “the creation of a thinking machine, which would be able to perfectly simulate the cognitive faculties of the human mind” (Natale and Ballatore 2020: 3). This problematic pattern is defined as “the recurrence of analogies and discursive shifts, by which ideas and concepts from other fields were employed to describe the functioning of AI technologies” (ibid.). Notably, the discourse on AI makes use of the cybernetic metalanguage (Agre (1997a: 134) describes this as ‘cyborg discourse’). Cybernetics, in turn, is strongly linked to the field of biology, treating machines and living beings as items on the same spectrum (‘systems’) but with different complexity.[23] Additionally,

[23] It often seems necessary to repeat, especially in the face of certain discourses on AI, that cybernetic metalanguage—or talking about humans and machines in similar terms—is a discursive manipulation. For semiotics, the heuristics in both cases are models; models (including discourse) are always simplifications, and the fact that we *can* use the same model as a heuristic to talk about features of living and non-living systems does not indicate the ontological sameness of the two research objects.

the AI myth was fuelled by the "rhetorical use of the future, imagining that present shortcomings and limitations will shortly be overcome," and the maintenance of controversies around the claims about AI (Natale and Ballatore 2023: 3). Coming back to the representationalist fallacy and the problem of confusing the map with the terrain, Johnson and Verdicchio argue similarly that "Although human autonomy may in certain contexts be a useful metaphor for the autonomy of computational artefacts, some scholars get caught up in the metaphor and seem to forget the difference between the thing and its metaphorical parallel" (2017: 585).

2.2.1 Discursive anthropomorphism

Anthropomorphism seems to be an inevitable part of human perception and sociocultural construction of the world. It "has been generally considered an invariant and automatic psychological process that is simply a chronic feature of human judgment" (Epley et al. 2007: 865). Anthropomorphism is most likely linked to our faculty of empathy, more precisely to cognitive empathy or perspective-taking. Social cognition studies show that people take a robot's visual perspective in a manner similar to perspective-taking toward other humans, increasingly so when the robot has human appearance (Zhao and Malle 2022). Empathy is also at play in social ethics and our underlying sense of ethical obligation towards other beings and things. Joanna J. Bryson and Philip Kime are convinced that the exaggerated fears and hopes concerning AI are the result of misplaced empathy, or an individual's "misidentification with machine intelligence" which "leads to false ethical evaluations of AI's potentials and threats" (2011: 1642). This misidentification is caused by "superficial, unfamiliar similarities such as a machine's use of language or reason"—that is, anthropomorphism—and can be mitigated by increased "empirical experience" with AI systems and "making AI visible and understood where it already exists" (Bryson and Kime 2011: 1645).

Anthropomorphism as a cognitive process has been discussed foremost in biology and theology. In both disciplines, anthropomorphic reasoning is typically considered pseudoscientific or poor science. However, in animal studies, some scholars argue that anthropomorphism could also be a useful heuristic when hypothesizing about non-human animal behaviour. Gordon Burghardt (1991) advocates for a critical anthropomorphism in ethology—to be able to gauge the intentionality of (allo)animals—as a heuristic method and an analogue to critical realism. However, he specifically cautions that "critical anthropomorphism is not useful in dealing with non-living entities" (Burghardt 1991: 75).

Primatologist Frans de Waal (1999: 262) distinguishes three kinds of anthropomorphism. Firstly, anthropocentric anthropomorphism is "a naive projection of human experience onto other animals" ("confusion between humans and animals"). Secondly, animal-centric anthropomorphism consists of more mature perspective-taking achievable for example via Uexküll's Umwelt concept ("understanding animals on their own terms"). Thirdly, there is heuristic anthropomorphism or the "identification with animals to develop testable ideas" (de Waal 1999: 262). The latter form also contains "mock anthropomorphism", where human features are attributed to animals in a playful way as part of the analytical process (de Waal 1999; Kennedy 1992). Mock anthropomorphism recognises that anthropomorphism is an unavoidable and useful part of human reasoning; that "we habitually anthropomorphize about animal behaviour, using our own mental processes as models to 'explain' the behaviour in terms of intentions" (Kennedy 1992: 89; de Waal 1999: 270). For de Waal, it is desirable to limit the methods of scientific inquiry to animal-centric and heuristic anthropomorphism and leave aside the anthropocentric kind of reasoning as much as possible, although the latter does have certain usefulness in modern culture (for example in literature or satire). With his reasoning, de Waal also attempts to overcome the still-pervasive Cartesian dualism and turn anthropomorphism into a useful heuristic model that could demonstrate that "animals are *not* automatons" (de Waal 1999: 271, original emphasis). Kennedy argues that anthropomorphic metalanguage may signify both genuine or mock anthropomorphism, with the author being possibly unaware of their own intentions while using anthropomorphic vocabulary (Kennedy 1992: 100).

Both de Waal and Kennedy mention cases where the object of anthropomorphism is an inanimate object—a machine, a computer: "We habitually resort to anthropomorphic metaphors also when we describe the complex behaviour of inanimate systems such as the weather or computers, but there is little danger that we shall take the analogies literally. This use of mock anthropomorphism runs no risk of being mistaken for genuine anthropomorphism" (Kennedy 1992: 99). De Waal suggests that mock anthropomorphism enters computer science conversations simply "because we find it easier to deal with machines in human terms" and not because we assume that "machines share our experiences" (1999: 270).

Previous research by my Tartu colleagues has shown that the representations of animals in popular culture, such as cartoon characters or toys, have an impact on human-alloanimal relationships. For example, the popularity of cartoons such as Teenage Mutant Ninja Turtles, Ratatouille and Finding Nemo have impacted the sales of pet animals – turtles, rats and clownfish respectively (Dydynski and Mäekivi 2021). Additionally, cartoon characters and toy animals can affect the perception of the real corresponding alloanimal species. Although the representations are clearly and distinctly fictional, they inevitably contribute to the formation of the corresponding "Expected Animal" representation of the real-world species as people internalize the character's aspects and project them onto the actual species (Dydynski and Mäekivi 2021: 759). In a similar manner, the science-fictional and public media representations of all things called "AI" contributes to the formation of an "Expected AI", which is therefore perceived as more autonomous, more independent, and more functional than current systems allow.

Mock anthropomorphism seems to have lost its heuristic self-awareness in certain discussions about robot rights or machine consciousness. Arguably, "treating robots like animals" is inappropriate and misleading (Johnson and Verdicchio 2018). And even while computer scientists have realised that endeavours in machine intelligence can also be inspired by the cognitive models of other-than-human life forms, this practice has side effects that discursively devalue non-human animals (Dydynski and Mäekivi 2021; Dydynski 2021) and reconstitute them as mindless Cartesian machines.

A growing body of literature addresses anthropomorphism in the context of technology and computer science. Anthropomorphism may be manifest in the appropriation of human characteristics to machines. "Experiment after experiment has shown that people are more likely to ascribe human qualities such as moral sensibility to machines on the basis of their humanoid appearance, natural language communication, or the mere fact of having been given a name" (Chesterman 2021: 127). One such feature is the autonomy that an observer attributes to or perceives from a working robot. The perceived autonomy of a robot tends to be higher when its appearance is more human-like (even if the 'appearance' is only virtual) (Harbers et al. 2017). The perceived autonomy also increases above average when the robot is operated from far away, it performs high complexity tasks or acts in a disobedient manner (ibid.). Therefore, the anthropomorphism of technology is largely motivated by the anthropomorphic design features which may be intentional.

In computer science, attitudes toward anthropomorphism are ambivalent. Anthropomorphic design in social robotics is proclaimed to increase usability and user-friendliness of the devices. De Visser et al (2016) have conducted an experimental psychological study to learn how the anthropomorphic features of a technology affect its adoption and acceptance by human users. They do not explicitly define anthropomorphism but treat it rather as a type of characteristic of the object (automated agent). Generally, anthropomorphic features are found to increase trust in the automated agent and the authors recommend anthropomorphism as a deliberate design choice (de Visser et al. 2016). However, there is also critique against anthropomorphism in social robotics, where it is seen as deceitful: "The potential harmful impact of deceptions that result in an overestimation of a robot's functionality include their inappropriate use to replace human care, and a misplaced trust in their ability to make decisions for which they are not qualified" (Sharkey and Sharkey 2020: 315).

The anthropomorphism of AI systems has been called the "android fallacy." Some (e.g. Richards and Smart 2016) caution against the "android fallacy" or the "tendency to anthropomorphize AI systems" (Chesterman 2021: 127). They stress the importance of choosing correct metaphors especially in legal context, even when it is commonplace to use human-specific nouns and verbs when describing the robots' parts or operations (Richards and Smart 2016: 18). From a military-legal perspective, Sigrid Johansen is firmly against anthropomorphism, stating that "autonomous weapon systems remain weapons and […] they do not become humans, although we use human-like characteristics to describe them. Furthermore, […] because we call these weapon systems autonomous, we attribute them with human-like behaviour that they are not likely to possess in the near future, and subsequently that this attribution of human-like behaviour is not beneficial for our relationship to the machines or to the assessment of legality" (Johansen 2018: 90).

Anthropomorphising rhetoric is ethically questionable and can be misleading when it comes to the public discourse on technologies. Watson (2019) contests the anthropomorphism of neural networks, because: "Finally, I shall argue that while the connections between machine learning algorithms and human cognition may be intriguing and suggestive, the rhetoric of anthropomorphism can do more harm than good when it comes to conceptualizing the important ethical challenges posed by emerging technologies" (Watson 2019: 418).

Anthropomorphism poses various legal problems because people tend to attribute the kind of agency to the systems that they do not possess (Johansen 2018; Chesterman 2021) and consequently presume that the responsibility or accountability for certain actions can be assigned to the machines as well. The debate over the possible moral responsibility of machines is a long, well-documented, and increasingly tedious waste of effort. For this reason, (Taddeo and Blanchard 2022: 78) consider it important to finally arrive at a consensus that "autonomous decision-making should not absolve humans from responsibility."

2.2.2 Automation bias

Multiple research threads concerning human-machine interaction have found that there are serious problems with employing automated workflows. Many of these issues stem from how machines, automation and AI are interpreted in culture. A major issue with delegating decision-making or any other cognitive shortcuts to machines is described as the problem of automation bias. Automation bias is defined as "the tendency of humans to use automation […] as a heuristic replacement for vigilant information seeking, cross-checking, and adequate processing supervision" (Johnson 2022: 444) or the tendency to attribute greater authority to the machine than one would to a human in a similar situation (de Visser et al. 2016; Cave and Dihal 2020). Multiple studies demonstrate how people tend to trust machine decisions over their own, simply because they believe that these are more objective (Mosier and Skitka 1999; Goddard et al. 2012; Kapania et al. 2022; Schemmer et al. 2022; Glickman and Sharot 2022). Automation bias is well-researched in the field of healthcare, where clinical decision support systems (DSS) are used to assist medics in diagnosing and processing information. For example, a study shows how "the occasional incorrect advice [the DSS] give may tempt [doctors] to reverse a correct decision they have already made, and thus introduce errors of overreliance," which happened in 6% of the cases (Goddard et al. 2012: 121). The overall amount of such errors may be small but even a small number of errors accompanied by automation bias might be unacceptable in the fields where a decision would have severe impact on someone's life or health.

Similar issues have been addressed under the concepts of *android fallacy* (Bryson and Kime 2011), *anthropomorphism fallacy,* or generally the tendency to anthropomorphise AI systems (Richards and Smart 2016). A separate body of social research addresses *algorithm(ic) appreciation* or *algorithm bias* (Araujo et al. 2020). A similar problem has been long discussed in aviation research as *(automation) complacency* (Wiener and Curry 1980; Wiener 1981; Parasuraman and Manzey 2010).

Recent research into people's inclination to trust machine "decisions" reveals even more worrying tendencies. A study published last year examines how human-AI interaction impacts and interferes with human decision-making (Glickman and Sharot 2022). In the experiment, participants were presented with a set of computer-generated images of human faces and asked to evaluate whether in their opinion, the face looked "more sad" or "more happy". The participants were then presented with an alternative evaluation that was framed as "AI response" and asked whether they wanted to change their answer or not. The study finds that the respondents "changed their response on 32.72% (± 2.3% SE) of the trials in which the AI provided a different response than they did, and on 0.3% (± 0.1% SE) of the trials in which the AI provided the same response than they" (Glickman and Sharot 2022: 6). The authors conclude that (biased) algorithms produce more biased humans. This is very significant for the majority of AI research happening right now, in the context of increasing amounts of research demonstrating the racial gender and socioeconomic biases present in datasets and consequent AI systems which severely amplify the bias (Hajian et al. 2016; Zhao et al. 2017; Buolamwini and Gebru 2018; Sap et al. 2019; Hassine and Neeman 2019; Raji et al. 2020; Keyes 2020; Johnson et al. 2022; Horowitz and Kahn 2023). The above are only few examples of studies revealing the severity of issues and looking for ways to achieve fairer AI systems.

The study elaborated above (Glickman and Sharot 2022), however, highlights a more dangerous trend in our use and understanding of AI: if over 30% of people are willing to change their opinion simply because "an AI said so", what are the implications for our societal institutions and processes that are converted to (semi-)automated decision-making systems? Machine learning systems are also characterized by fundamental epistemic opacities, meaning that it is not possible to explicate the machine decision-making process to an extent that we could understand what and why a certain decision was made (Grünke 2020; Héder 2023). While this opacity might not be a problem in applications such as games, the situation becomes very different when we are developing systems that have direct impact upon and the power to change our social structures and institutions.[24] For example, the use of biometrics in law enforcement or security systems (Santosh and Wall 2022) and predictive policing (Berk 2021) can have critical impact on the people subjected to them. Another level of issues rises with what Nancy Leveson calls safety-critical systems such as medical devices, weapons systems, aircraft, or nuclear power plants (Leveson 2017). She argues that such systems should be subjected to much higher safety requirements than other

[24] Timnit Gebru argues that "we have to be very explicit […] about what our error rates are."—Rony Chow, "Timnit Gebru: The Computer Scientist Fighting for a Fairer World." *History of Data Science*. August 19, 2021. https://www.historyofdatascience.com/timnit-gebru-the-computer-scientist-fighting-for-a-fairer-world/.

software, and especially that safety-critical systems should not by any means be allowed to operate on probabilistic prediction mechanisms (ibid).

Algorithmic decision-making is especially problematic when it comes to automating workflows in public services, credit scoring, and policing, to name a few. Various studies have shown that people are often over-targeted because of their belonging to a certain social, racial or socioeconomical group. Members of marginalized groups face higher levels of scrutiny and are often at an additional disadvantage when their behaviour is red-flagged by an algorithm and they have to contest the automated decision (Eubanks 2017).

Often, the problem arises not from the biased algorithmic judgment per se, but from the underlying design of the system or even the preceding problem formulation. Deborah Raji and colleagues (2022) list a taxonomy of AI functionalities that can lead to possible failures, as well as an extensive number of historical examples. Sometimes the entire premise of an AI system is built on outdated pseudoscience. For example, phrenology—or the idea that it is possible to deduce emotions, intentions, or other characteristics from a person's facial features—is foundational for most of the "emotion recognition" software, as well as some predictive policing tools.

Eventually, as Gary Smith argues in *The AI Delusion*, "the real danger is not that computers are smarter than us, but that we *think* computers are smarter than us and therefore trust computers to make important decisions for us." Therefore, "We should not be intimidated into thinking that computers are infallible, that data mining is knowledge discovery, [or] black boxes should be trusted." Eventually, "Statistical evidence is not sufficient to distinguish between real knowledge and bogus knowledge" (Smith 2018: 237).

2.3 Who can talk about technology?

One of the most typical presuppositions that comes up again and again in relation to technology in general and AI in particular is the claim that technology is neutral—that its builders bear no responsibility for its use, and that it is the end-users only who choose to use a given technological tool for good or evil ends. Proponents of the neutrality thesis argue that technology is value-neutral, a passive tool (Oberdiek 1990). This instrumentalist perspective characterises a large part of popular discourse on AI, especially among Silicon Valley tech elite.[25] The neutrality thesis is, for instance, underlying the choice of making text and image

[25] For example (Atkinson et al. 2019) dedicate a long report characterising technology critique as "techlash", claiming it is unfounded and unfair against technology companies which, among other things, far from spying on unsuspecting people, "are clear about their growing use and reliance on data" while "most consumers happily accept free services knowing full well they are providing data" and are, in the authors' opinion, completely free to use alternative services if they care about privacy (ibid, 9). The datafication of the semiosphere and its impact on culture and society is, again, another avenue for future research.

generators such as ChatGPT, DALL-E and Midjourney available to the public, while the ability of the given tools to produce racist content and hate speech was widely known or at least not ruled out. Neutrality thinking assumes that the responsibility for the use of the model lies with the people—its end-users, and the producer-company such as OpenAI cannot be held accountable for the use of the tool. Among others, Oberdiek traces the neutrality argument to Melvin Kranzberg's apologetic defence of nuclear technology in the 1960ies (Oberdiek 1990: 73, 75). Kranzberg, the much-cited historian of technology, is also the author of the famous "Kranzberg's Laws" which outline the relationships between technology, people, and society. In these laws—describing "the development of technology and its interactions with sociocultural change," formulated by 1985 and printed in 1986—Kranzberg takes a more critical stance towards technology and its impact on society. Thus, the first statement argues: "Technology is neither good nor bad; nor is it neutral" (Kranzberg 1986: 545) Further, he explains:

> "By that I mean that technology's interaction with the social ecology is such that technical developments frequently have environmental, social, and human consequences that go far beyond the immediate purposes of the technical devices and practices themselves, and the same technology can have quite different results when introduced into different contexts or under different circumstances." (Kranzberg 1986: 545–46)

Kranzberg's contribution to understanding technology as a socio-cultural construct and result of power structures is a topic that certainly deserves further study but remains out of the scope of this thesis.

The idea of techno-neutrality has not subsided and needs to be contested every now and then, thus there are many contemporary approaches that still contend with it (Winner 1980, 1997; Oberdiek 1990; Hare 2022). Most recently, Luciano Floridi (2023) argues that the design of any technology is a moral act, hence the ones responsible for the design, engineering, and construction of any AI system cannot hide behind the neutrality thesis.

Stephanie Hare argues that "technology is too important to be left to technologists" and that "To ignore technology is a decision—one that turns us into a cog in someone else's machine" (Hare 2022: 6–7). Participating in the discourse on technology is, however, not that simple. The general securitization of the discourse, as understood by the Copenhagen school, is an effective manoeuvre to delimit who has access and right to participate in the discourse on AI in the first place. Unfortunately, the discourse on AI is often securitized, especially when it comes to Apocalyptic AI which is one of the favourite "security" arguments from certain technologists. For the Copenhagen School, "security is a speech act that *securitizes*, that is constitutes one or more referent objects, historically the nation or the state, as threatened to their physical or ideational survival and therefore in urgent need of protection" (Hansen and Nissenbaum 2009: 1156, original emphasis). In the case of AI, the threatened referent object is the whole of humanity, abstracted into a homogenous mass, to whom the bright minds of Silicon Valley hasten to provide the desperately needed service of preventing the apocalypse.

When it comes to the internal working logic of a computer system—what is characterised as a "black box" by Ashby and others—Martin Irvine argues that this, too, serves rather as a political device for intentionally rearranging power relations between the designers of the system and the "'users' as passive consumers". It is not just limited to the technical "unknowability" or what was earlier characterised as the epistemic opacity of AI systems.

> "The conceptual metaphor 'black box' was originally an engineering term for any component designed to take in certain kinds of inputs (energy, signals, information, etc.) and convert them into specified outputs (e.g., a radio, a voltage transformer, a codec for converting digital into analog audio/video): the details inside the components can just stay 'hidden' ('black-boxed', 'don't need to know', 'built-in'), because only the outputs matter for the design purpose. The concept is now universally used in systems and software design: 'blackboxing' is used to hide the internal complexities of a module (at one functional level) that other modules in a system 'don't need to know about' for using the outputs communicated to system (see below on systems and design theory). But in our contemporary political economy for intellectual-property-protected products, this design principle is also used intentionally to close off access to computing systems in 'black-boxed' manufactured devices, which are intended *to maintain 'users' as passive consumers blocked from understanding the universal semiotic principles on which the devices depend*." (Irvine 2022: 205–6, added emphasis)

Following Irvine's argument, AI designers are not just technically unable to explain the inner workings of a system; they are not motivated to provide any explanations at all. Nowhere is this implicit intent more apparent than in the arguments about how AI is "a real thinking machine" or at least on the path to one. Drawing on the taken-for-granted status offered by societal power structures, these black-boxed imposed relevance structures can act as what Ritivoi (2009: 36) calls *repositories of normativity*, indicating the legitimised and normalised knowledge items in a given society. Such repositories of relevance always contain power structures (Dreher 2016) that are guiding the normalization of certain attitudes. The power problem is manifest in the question of *who is allowed to talk about technology at all*. A large part of public AI discourse entails citing various "experts" and arguing over which ones we should listen to. All the while, "criticisms of the field, no matter how sophisticated and scholarly they might be, are certain to be met with the assertion that the author simply fails to understand a basic point" (Agre 1997a: 132). In the constructivist terminology of the Copenhagen school, the power plays in science and technology are formalised in the *technification of the discourse* which grants a privileged role to computer scientists and engineers as the only experts "who have the authority to speak about the unknown" (Hansen and Nissenbaum 2009: 1166–67). Incidentally, in a situation where "the science is highly uncertain, expert evaluations are more likely to reflect the background assumptions, priorities, values, and imagination of the experts"

(Pamuk 2021: 10). Accordingly, scientists in AI discourse frequently use examples from science fiction to describe both current technologies and anticipated future developments; this happens both in popular scientific content (such as popular documentaries) as well as in academic books and articles.

Interestingly, Steve Woolgar (1985) draws attention to a very similar tactic pertaining to the discourse on expert systems with its implications for the sociology and philosophy of science. Having "mobilised the distinction between man and machine in claiming their own particular (human) expertise to speak about expert systems (machines)," "certain individuals"—practitioners, engineers, computer scientists—"define the nature and character of the object of study," all the while "claim[ing] to be uniquely competent in speaking on behalf of these objects" (Woolgar 1985: 565–66). It seems an understatement to say that for almost 40 years, the general discourse on AI has not changed at all. The "practitioners'" recent commentary regarding the deployment and development of large language models—kinds of "expert systems" in their own right—is straight out of Woolgar's textbook.

The *neutrality thesis* has also begotten the term *dual-use* technology. Dual-use usually indicates one of the following two ideas. First, it can be a technology used both for military and civilian purposes. Thus, we can speak about nuclear fusion as dual-use, for example, or drones. Certain unmanned vehicles could be used for both military and civilian purposes.[26] Secondly, dual-use can imply the perceived *good* or *evil* use of a technology. Here, civilian and military purposes are sometimes taken to roughly correspond to 'good' and 'evil' but not entirely. The use of nuclear fusion can easily be imagined in the categories of civilian as good (producing energy, electricity) versus military as evil (nuclear bombs kill people). With other technologies, the distinction is not so clear-cut—knives can be used in civilian context both for good (cutting food) and evil (cutting people) ends. While the original distinction of dual-use arguably concerns the opposition of civil and military spheres, public media discourse tends to conflate this good-evil distinction and blur the boundaries between all categories.

Steering the discourse toward expert-only decision-making and the sense that certain technological developments are inevitable is beneficial to the technologists for yet another reason. Wiebe Bijker argues that "Determinism inhibits the development of democratic controls on technology because it suggests that all interventions are futile. This is as true for science as it is for technology" (Bijker 1995: 281). Thus, the threat to democracy is not so much in the fear of proliferation of deepfakes, fake news and the consequent "epistemic apocalypse" (Habgood-

[26] THeMIS, an unmanned ground vehicle (UGV) developed by Estonian company Milrem is, for example, a vehicle that, with various upgrades, can be used as logistics, combat, or reconnaissance support for the military, but can also be repurposed and produced as a multi-scope commercial UGV to support fire brigade, search and rescue missions, or various other commercial applications that necessitate similar functions, such as last mile support or access to areas and situations that are dangerous for human workers.—https://milremrobotics.com/defence/, https://milremrobotics.com/commercial/. Accessed September 8, 2023.

Coote 2023). The threat is embedded rather in current, ongoing policy discussions, decision-making, regulations development, and other everyday contexts of existing sociocultural institutions which AI corporations try to tweak in their favour with their lobbying. Delegating part of the critical discussion about the impending changes of sociocultural systems in the hands of self-proclaimed 'experts' is only one part of a functional strategy. Overlooking all current, ongoing harms of these technologies to society, culture, and climate, and declaring the threat of an impending AI apocalypse more important than anything else forms another move. The fact that the discursive strategies have morphed into a self-perpetuating cult—as many of the researchers genuinely believe in said apocalypse[27] or argue vehemently that "machines are also people"—is in part ingenious, in part terrifying.

The anti-democratic tendency is not particularly new either. "A recent study of American business leaders," Langdon Winner reminds us (1980: 133), "found them remarkably impatient with such democratic scruples as 'one man, one vote'. If democracy doesn't work for the firm, the most critical institution in all of society, American executives ask, how well can it be expected to work for the government of a nation—particularly when that government attempts to interfere with the achievements of the firm?" The cited authors Leonard Silk and David Vogel conclude that for the businessman, "the pattern of authority and reward within the firm becomes [...] the desirable model against which to compare political and economic relationships in the rest of society." (Silk and Vogel 1976: 191; see also Winner 1980: 133).

Accordingly, it is not uncommon for technologists to pressure governments and legislators toward decisions favourable to their corporations. The public discussions that took place over the summer of 2023 around the EU AI Act provide a fitting example. In May 2023, the CEO of OpenAI indicated that the company might "cease operating" in the EU if the AI Act in development classifies its systems as "high risk."[28] At the same time, OpenAI lobbied for significant elements of the AI Act "to be watered down in ways that would reduce the regulatory burden on the company," achieving desirable changes of wording in the final draft of the Act that was approved by the European Parliament on June 14, 2023.[29] Nevertheless, the principal problem in the adoption of policies seems to lie in the semiotics—the categories and the definitions of terms. An earlier report on the developing Act criticizes its treatment of both the concepts of AI and "high risk"

[27] A recent story in WIRED reports that most OpenAI employees believe in AGI; or at least, the executives expect them to: "It's not fair to call OpenAI a cult, but when I asked several of the company's top brass if someone could comfortably work there if they didn't believe AGI was truly coming [...] most executives didn't think so. *Why would a nonbeliever want to work here*? they wondered. The assumption is that the workforce [...] has self-selected to include only the faithful."—Steven Levy, "What OpenAI Really Wants." *Wired*, September 5, 2023. https://www.wired.com/story/what-openai-really-wants/. Accessed October 1, 2023.

[28] Billy Perrigo, "OpenAI Could Quit Europe Over New AI Rules, CEO Warns," *Time*, May 24, 2023. https://time.com/6282325/sam-altman-openai-eu/. Accessed October 15, 2023.

[29] Billy Perrigo, "Exclusive: OpenAI Lobbied E.U. to Water Down AI Regulation." *Time*, June 20, 2023. https://time.com/6288245/openai-eu-lobbying-ai-act/. Accessed October 1, 2023.

(Grady 2023). Following the lobbied changes adopted in the last draft of the Act,[30] the European Digital Rights (EDRi) association called for closing the "dangerous loophole" that would allow the AI companies themselves—and not a separate party—to decide whether they want to designate their products as "high risk" or not.[31] As of 19 October 2023, EU Parliament's legal office was reported to have condemned the introduced possibility for the developers of AI to decide their own filter conditions, noting that it would be entirely in contrast with the regulatory purpose of the AI Act.[32]

How, then, should we approach the discourse on AI to reappropriate our human decision-making rights? Yarden Katz argues for completely "refusing AI—the concept, the associated epistemic forgeries and models of the self, as well as the institutions that sustain these" (Katz 2020: 234). Yet, he does not advocate the complete rejection of technology, only the ideology of AI as it is now, and which should be rebuilt in different ways and empower different agencies:

> "Refusing AI is only part of refusing whiteness, as an ideology and institutional logic, while recognizing that such acts can be realized only collectively. Refusal is not the end, just the continuation of something different." (Katz 2020: 234)

Eventually, what we might be experiencing is simply "the end of a history that fabricated the ontology of the present, the slow disintegration of the [colonial matrix of power]" (Mignolo 2021: xxi).

---

30 The EU document P9_TA(2023)0236 detailing the amendments adopted on 14 June 2023 can be found on https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html. Accessed October 1, 2023.

31 "EU Legislators Must Close Dangerous Loophole and Protect Human Rights in the AI Act." *European Digital Rights (EDRi)*. September 7, 2023. https://edri.org/our-work/civil-society-statement-eu-close-loophole-article-6-ai-act-tech-lobby/. Accessed October 1, 2023.

32 Luca Bertuzzi, "AI Act: EU Parliament's Legal Office Gives Damning Opinion on High-Risk Classification 'Filters.'" *Euractiv*. October 19, 2023. https://www.euractiv.com/section/artificial-intelligence/news/ai-act-eu-parliaments-legal-office-gives-damning-opinion-on-high-risk-classification-filters/. Accessed October 20, 2023.

Today's tech barons will tell you that they uphold international human rights norms and that their particular supply chains are clean. They will assure you that conditions are not as bad as they seem and that they are bringing commerce, wages, education, and development to the poorest people of Africa ("saving" them). They will also assure you that they have implemented changes to remedy the problems on the ground, at least at the mines from which they say they buy cobalt.
After all, who is going to go all the way to the Congo and prove otherwise, and even if they did, who would believe them?

—Siddharth Kara,
Cobalt Red: How the Blood of the Congo Powers Our Lives

2.4 Toward a more ethical ontology of AI

Certain technologies—such as computers, information systems, and AI—are considered ubiquitous (Agre 1997a) or umbrella technologies (Horowitz 2019), in a similar fashion as electricity and the telephone. Ubiquitous technologies and their applications are broad and used in so many different fields of practice that it may be very difficult, if not impossible, to understand their precise effects on society (Agre 1997a). Therefore, specific efforts must be made to study technologies and their effects in context, as embedded in cultures and social practices.

The editors of *AI: the Case Against* argue that "the intelligent use of computers in artificial information processing contains a strong ethical component which includes a scientist's responsibility for the influence of her/his results upon the development of mankind" (Born and Born-Lechleitner 1987: ix). Eventually, "Responsible computing ultimately requires that technical communities develop and adopt tools, processes, and practices that mitigate harms and support human lourishing" (McMillan-Major et al. 2023: 1). Bender et al. (2021) outline various risks and problems with currently popular large language models, including high environmental and financial costs; various problems and biases in training data; the incapability of data to accurately represent certain sociocultural realities; the semiotic problems with understanding the limit and scope of language technologies; and accurate discursive interpretation of the sociocultural role of such technologies.

Power and politics are at the heart of many problems associated with contemporary AI technologies. The politics of AI highlights an important issue that cannot be overstated: that almost all robots and computer systems ("any artificial entity situated in the real world that transforms perception into action"), besides affecting communication and information processes, "produce direct physical impact on the world" (Bryson 2010). This impact penetrates in some manner almost all aspects of our social lives and organisations, cultures, economies, politics, and the environment. Various recent extensive studies attempt to gauge the impact of AI for business (Appio et al. 2023), finance (Kahyaoğlu 2021), and intellectual property law (Lee et al. 2021). When developing a specific system, the concerns and impact on various facets of global and local lives need to be accounted for. Johnson and Verdicchio (2017) argue for a new, expanded ontology for AI that "will allow ethical issues to be more readily seen and addressed" (Johnson and Verdicchio 2017: 577). They propose the concept of *sociotechnical ensembles* as "combinations of artefacts, human behaviour, social arrangements and meaning" (ibid: 583). Above all, this approach makes it possible to clearly communicate that "putting into the world a robot that has the capability of harming humans is a *human act*, and the human actors who release such computational artefacts will be responsible for the consequences, not the computational artefact itself" (ibid: 589, added emphasis). This framework enables shifting the focus "from futuristic computational artefacts to those who design and build them and embed them in social contexts" (ibid: 589). This is already a good step for going beyond AI as objects and starting to outline the complex impact networks that they have.

However, I would like to go even further.

The underlying geopolitics of knowledge (Mignolo 2011) is what makes AI especially impactful on a global scale. As already observed for modernity, the Western paradigm for sociocultural existence relies on its dark side—a specific colonial epistemology, "knowledge-making and the global power differential" (Mignolo 2011: 140). In this modern/colonial world order,

> "[…] the racialization of places and people in the formation and transformation of the colonial matrix of power not only established hierarchical ranking between languages and categories of thought, but also built economic and political structures of domination and oppression based on the geopolitical and hierarchical organization of knowledge." (Mignolo 2011: 141)

Therefore, an ethical ontology of AI needs to be able to consider the problems with the underlying world system itself—the underlying social and political order based on imperial colonialism and oppression of the Global South. This system, also discussed as the colonial matrix of power (Mignolo 2011; Mignolo and Walsh 2018), European universalism (Wallerstein 2006), global white supremacy (Mills 1997) or White Crisis (Ali 2017, 2019), finds further objectivation, legitimisation, and crystallisation in and through most of the AI systems that we have today.

Western-European thinking that underlies our hegemonic systems of knowledge is thus permeated not only by racial capitalism (Wallerstein 1974, 2004) but also by racial epistemology (Mignolo 2002). Eventually, "What geopolitics of knowledge is unveiling is the epistemic privilege of the First World. In the three worlds of distribution of scientific labor, the First World had indeed the privilege of inventing the classification and being part of it" (Mignolo 2011: 129). For Wallerstein, the "links between universalism and racism [serve] as justifications for the exploitation of labor" (Mignolo 2002: 78). The old capitalism relied on explicit slavery—the racialisation of people, the commodification of everything, and the control of the living (Mbembe 2019: 177). The new neoliberal capitalism entails "the appearance of an ever-growing class of slaves without masters and masters without slaves" where "self-reification constitutes the best chance of self-capitalization" (Mbembe 2019: 179, 178).

Platform owners argue that people are entirely free to choose whether to participate in these extractive global sociotechnical networks. They assert that people receive fair compensation for their labour and data, even when it only amounts to the possibility to use the platform "for free". Unfortunately, the global scale of surveillance, microtargeting, underpaid click-workers[33], and fauxtomation[34] reveals a different world. Instead, the new surveillance capitalists "insist upon the 'freedom to' launch every novel practice while aggressively asserting the necessity of their 'freedom from' law and regulation" (Zuboff 2019: 495). The earlier example of attempts to influence the EU AI Act illustrates this quite well. Instead of free and equal benefits for all people we are moving deeper into the world characterised by various forms of modern slavery (cf. Kara 2014, 2017, 2023).

The overwhelming carbon impact of AI models is becoming increasingly evident (Dhar 2020; Nat 2023). Many studies try to assess the carbon footprint of training various AI models (Lacoste et al. 2019; Ligozat et al. 2022; Luccioni et al. 2022; Li et al. 2023) or gauge the footprint of the global information and communication technology (ICT) sector in general (Belkhir and Elmeligi 2018; Wang and Xu 2021; Bieser et al. 2023). The estimate for training a large Transformer model such as GPT-2 is about 280 000 kg CO2 (626 155 lbs) (Strubell et al. 2019)[35]. A 2018 calculation by OpenAI researchers shows that, while the "the amount of compute used in the largest AI training runs" has doubled every two years in the period 1958–2012 (since Rosenblatt's Perceptron), it has doubled every 3.4 months since 2012, leading to a 300 000x growth in the period 2012–2018.[36] Sasha Luccioni points out that for the most recent LLMs such as GPT-4 we do not know and cannot estimate their footprints because OpenAI has not made public the necessary data for that.[37] But the carbon imprint of AI training is only part of the problem—the carbon impact of the entire technological infrastructure upholding AI is even more significant (Dhar 2020: 424). ICT sector was estimated to contribute around 2–4% of total global emissions in 2020 (Freitag et al. 2021). However, the assessments lack a unified method and employ sporadic amounts and types of data. Consequently, the studies rely on different sets of technology categories and vary widely in how they allocate responsibility for environmental impact across subsectors. For instance, Freitag et al. (2021: 5) review studies estimating the carbon share of data centres anywhere from 18% to

---

[33] For example, content moderation, as well as data labelling—necessary to train any AI system—are tedious jobs that companies prefer to outsource to "call centres" with poor pay and working conditions. — Billy Perrigo, "Inside Facebook's African Sweatshop." *Time*, February 14, 2022. https://time.com/6147458/facebook-africa-content-moderation-employee-treatment/. Accessed October 15, 2023; Billy Perrigo, "OpenAI Used Kenyan Workers on Less Than $2 Per Hour: Exclusive | Time." *Time*, January 18, 2023. https://time.com/6247678/openai-chatgpt-kenya-workers/. Accessed October 15, 2023.

[34] *Fauxtomation*—a term coined by Astra Taylor (2018)—refers to the practice of portraying certain work processes as automated while the work is actually done by people.

[35] GPT-2 by OpenAI was launched in 2018. The same study inspired media headlines and articles of the AI training requiring the five lifetimes of CO2 emissions of a car.

[36] Dario Amodei and Danny Hernandez, "AI and Compute." *OpenAI*. May 16, 2018. https://openai.com/research/ai-and-compute. Accessed September 8, 2023.

[37] Sasha Luccioni, "The Mounting Human and Environmental Costs of Generative AI." *Ars Technica*, April 12, 2023. https://arstechnica.com/gadgets/2023/04/generative-ai-is-cool-but-lets-not-forget-its-human-and-environmental-costs/. Accessed October 1, 2023.

41% and consumer devices from 32% to 57%. Most authors agree that the energy consumption of the ICT sector is rising quite dramatically. An earlier study predicts that in the worst-case scenario, by 2023, communication technologies could account for as much as 51% of global electricity consumption and 23% of the greenhouse gas emission—that is, if not curbed by the concurrently rising financial costs (Andrae and Edler 2015).

There are many studies focusing specifically on “Sustainable AI”. But this concept has been challenged as an empty term and greenwashing because many cases simply seek to build predictive models loosely associated with UN sustainability goals, and do not count for their real efficacy nor the carbon impact of building the models themselves (Falk and van Wynsberghe 2023). Aimee van Wynsberghe argues for a clear distinction between AI *for* sustainability, and Sustainable AI as “a movement to foster change in the entire lifecycle of AI products […] towards greater ecological integrity and social justice” (van Wynsberghe 2021: 213).

Most studies of the social impact of technology typically focus on the post-production stage, involving only the reception of AI in society. Current models are looking at the period after a technological tool has been made publicly available and is adopted by social groups, cultures, or whole societies. At most, various HCI (Human-Computer Interaction) models include the immediate designers of a given computer system—the kinds of actors visualised in Figure 1.

No AI could exist without the vast ICT infrastructure embedded in our planet; in turn, this infrastructure, the instances of AI and their power consumption requirements have a vast impact on the environment, particularly regarding the transformations occurring due to climate change. Thus, AI can be observed as a much bigger figuration that includes the entire production chain and possibly even more. Kate Crawford argues that “the core issue is the deep entanglement of technology, capital, and power, of which AI is the latest manifestation,” creating “the planetary infrastructure of AI as an extractive industry: from its material genesis to the political economy of its operations to the discourses that support its aura of immateriality and inevitability” (Crawford 2021: 217). As seen in chapter 2.3, this sense of inevitability is supported by various discursive mechanisms and technological-deterministic arguments. There is also need for models that account for the complexity and interconnectedness of contemporary socio-technical assemblages and their power relations which extend beyond states and legislations, our habitual units of social self-organisation.

Figure 5 visualises a selection of discursive and ontological aspects that have become apparent during my work for this thesis. Far from being extensive, these three are some initial categories of AI impact and interaction with our world that are not receiving enough attention right now. Disregarding them further may, however, lead to potentially dangerous consequences for the entire planet, as visualised in the “collapse circles”. I will try to briefly explain these three facets below; however, they strongly resist separation from each other. In every example, these three aspects are intertwined and the impact of an action or a choice in the AI value chain runs in all directions. The labour of thousands or even millions of people is strongly implicit in each aspect and cannot be disregarded; the same goes for the interconnected impact on the lives of people, social groups, and so forth.

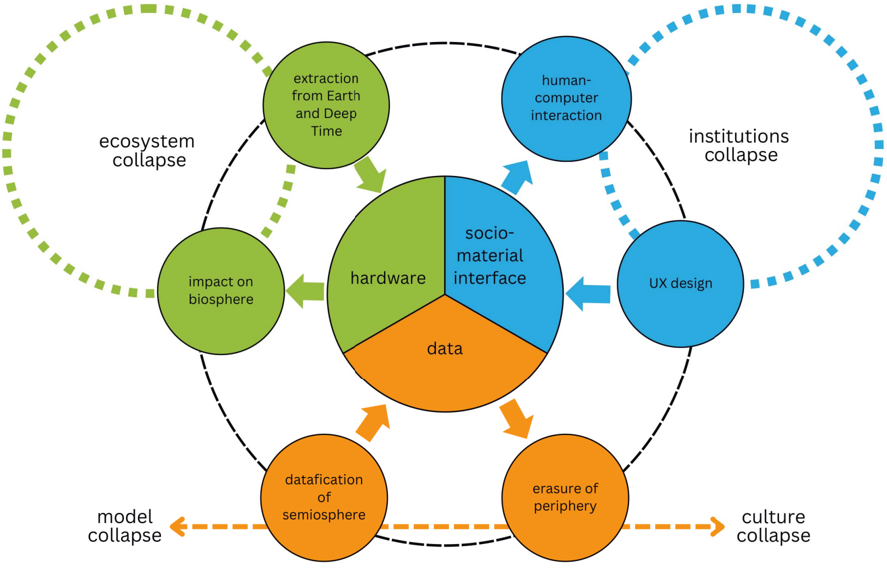

Figure 5. AI systems impact culture, society, and environment via their sociomaterial relationships. Reaching the system's tolerance limit, these relationships can lead to various collapses. This is a very initial model of what I call the Ecologies of AI.

As discussed in chapter 2, AI is a socio*material* object. Its materiality includes various strands of *hardware*, without which the said system would not exist. Additionally, as visualised by Crawford and Joler (2018), an AI system such as Amazon Echo utilizes large amounts of *data*, which is circulating in the system. Both the hardware and data come from somewhere, they are inevitable for AI, and they are the result of the work by thousands of people across the globe. Every single one of these people comes with their own sociocultural embedding, relationships, practices, and is in turn impacted by various strands of the global industry that is AI. AI is not just a problem of the moving object and its impact on its end users; AI is the problem of its entire production chain.

It can be argued that AI systems do not qualitatively differ that much from earlier computer products and traditional economics has never assigned much consideration to the human and environmental cost entailed in the entire production chain. Even while the field of environmental economics has since long attempted to tackle these problems through the concept of *externalities*[38], it is still far from

[38] Externalities are defined as "conditions arising when the actions of some individuals have direct (negative or positive) effects on the welfare or *utility* of other individuals, none of whom have direct control over that activity. In other words, externalities are incidental benefits or costs to others for whom they are not specifically intended" (Hussen 2004: 54, original emphasis).

seeing the people and their individual well-being in the value chain. So why AI, why now?

Firstly, I think it is a grave mistake (one that we will likely pay for with our lives in the foreseeable climatic landscape of the Earth) to disregard the full product chains and their impact in general. In the recent decade, there have been various attempts at mapping the "full life cycle" of products, which have resulted in practices and eco-labels such as Fairtrade etc. But this is far from being enough.

While the Western world is discursively racing toward a more "ecological" economy, "our quest for a more ecological growth model has resulted in intensified mining of the Earth's crust to extract the core ingredient—rare metals—with an environmental impact that could prove far more severe than that of oil extraction" (Pitron 2023: 9). It is not at all certain that developing more or different technologies would lead to better ecologies. Rather the opposite—our race for inventing new technologies—will most likely lead to new levels of problems. As a response to the technocratic visions, French engineer Philippe Bihouix advocates looking for more robust solutions in *low tech* instead (Bihouix 2014, 2020).

Secondly, AI is a uniquely global industry. The Semiconductor Industry Association (SIA) already claimed that about the semiconductors which make up only a tiny part of the hardware needed for AI (SIA 2016). AI systems are utilizing globally available knowledge, data, labour, and infrastructures; AI has a global impact in how it functions. Crawford and Joler (2018) argue that all the resources that AI systems use (as hardware+software+data) are part of the Commons, or our global common non-renewable resources. These resources are exploited and extracted from Earth and semiosphere by companies for profit. Therefore, AI systems, more than anything else before, need to be described as the totality of their sociotechnical construction and impact. For Crawford and Joler, this totality is composed of two streams of resources: the material stream of AI starts with mining rare earth metals; and the data stream starts with the quantification of culture—or the datafication of the semiosphere.

How, then, to approach this complex intertwined system of people, cultures, objects, materials, interdependencies, product chains, local and global economies? Sociologists Nick Couldry and Andreas Hepp (2017: 57), arguing that the most popular concepts so far—networks and assemblage—are "inadequate for understanding the distinctive processes of institutionalization characteristic of social life under conditions of deep mediatization". They suggest a new concept *figurations of figurations*, derived and developed further from Norbert Elias' social theory. For Elias, the concept of *figuration* highlights the role of people in societies and networks, without reducing them to isolated Weberian 'ideal types' (Elias 1978). Increasingly popular, 'network' is a problematic concept because "it reduces the social world to *nothing more than* the actor-constellations of networks" and thus "network theorists are unable to integrate [the] processes of meaning […] that *orientate* human action" (Couldry and Hepp 2017: 61, original emphasis). The figuration, on the contrary, enables the conceptualisation of both the individual and the society not in mutual exclusion and opposition, but as interwoven and interdependent (Elias 1978).

Datafication[39] *of the semiosphere* and platform economies,[40] which anything "AI" is just another instance of, remake the social world of various social communities into truly miserable ones without the people in these communities having any agency at all in deciding what kind of social order they want to have. This is the reversion and perversion of democracy, as criticized by Silk and Vogel (1976), on a whole new level.

The hardware section can be now best illustrated with an example of cobalt, recently discussed by Siddharth Kara in his book *Cobalt Red: How the Blood of the Congo Powers Our Lives*. Cobalt is a metal essential for rechargeable batteries, making them safer and lasting longer. Kara summarizes the importance of cobalt for the US and the EU as follows:

> "Cobalt is used in the manufacture of superalloys for turbines and jet engines; as a catalyst for cleaner fuels; in carbides used to make cutting tools; in materials used for dental and bone surgeries; in chemotherapies; and in the cathodes of rechargeable batteries. Given its wide range of uses, the European Union has designated cobalt to be one of twenty "critical" metals and minerals, and the United States has designated cobalt to be a "strategic mineral." (Kara 2023: 23–24)

Given the declared strategic importance of cobalt, the political designations almost sound like a justification for acquiring it "by any means necessary". The figurations of figurations that are global corporations with their hyperglobal value chains certainly push and pull towards this justification. The situation, placed in the context of the living conditions of an artisan cobalt miner as part of that 'value chain', is beyond horrifying. So how can we rectify it? Can it be done at all? What steps do we need to take, as Western societies, to move towards righting this wrong? It seems that whole economic paradigms need to be changed, but where to begin? Through their product and value chains, AI systems are inextricably linked to the issues of capitalism and climate change. In the global context, "The climate crisis is a crisis of racial colonial capitalism" (Simpson and Pizarro Choy 2023: 1), and so too AI cannot be seen as detached from racial capitalism. Foremost, we need to ensure the survival and flourishing of peripheral[41] thought, currently under increased threat from AI systems reproducing colonial Eurocentric patterns. There are research perspectives offered by myriads of brilliant scientists and activists in all overexploited and highly effected communities, to whom we should listen. Many decolonial scholars offer various points of entry into these bright and diverse worlds of thought (Mignolo and Escobar 2010; Gómez-Barris 2017; Cadena and Blaser 2018; Ndlovu-Gatsheni 2018; Smith 2021, to name a few). Or, as

---

[39] Couldry and Hepp define datafication as a form of objectivation, which due to its pervasiveness generates "new forms of coordinated social action" and "new forms of inequality regarding time and other resources" (Couldry and Hepp 2017: 139–40, 193).

[40] Shoshana Zuboff has written extensively about the social impact of automation (1988) and platform economies (2019).

[41] Peripheral in the Lotmanian sense. Laura Gherlone (2023) discusses the relevance of Lotman's work to decolonial scholarship and overcoming binary polarization.

suggested by Zoe Todd, we can start "by citing and quoting Indigenous thinkers directly, unambiguously and generously" (Todd 2016: 7).

In Western media, AI is often portrayed as the magical solution that delivers us from apocalypse or simply a means to arrive at a greater, better society on a global scale. This vision, however, benefits only the privileged few in limited geographical areas. For most of the world, AI entails new objectivations of colonial power and forms of intensified exploitation, and for the Congolese artisan miner, it gains the meaning of endless suffering in modern slavery in the name of the rechargeable devices available to unfathomable foreigners.

The personified depiction of AI as some kind of entity in potential need of "human rights" is mistaken and dangerous for many reasons. Most of all, such representation devalues and dehumanizes *all* people working anywhere in the global AI product chain who are not privileged enough to have their names listed in the model release paper[42]. At the same time, no AI would exist without its materiality, and without the perpetually collected and processed semiotic data. Looking at AI as entities or objects effectively erases from history the labour contributed by millions of people, and in longer perspective, their entire existence. After the centuries of colonialism that underlies the Western economy and makes it remotely possible at all to have these kinds of technologies today, the continuation of such practices cannot be justified.

Also, an ethical quest to "save the world" would focus on the well-being of each individual human being present in the current AI production chain, and not on the imaginary digital apotheosis of the posthuman TESCREAL life.

[42] Such as the release papers for GPT-2 (Radford et al. 2019), GPT-3 (Brown et al. 2020) or GPT-4 (OpenAI 2023).

3. CONCLUSION AND DIRECTIONS FOR FURTHER RESEARCH

As we parted ways, Augustin had this to say, "Please tell the people in your country, a child in the Congo dies every day so that they can plug in their phones."

—Siddharth Kara, Cobalt Red

This thesis explained and illustrated some ways in which the representations and ontologies of AI are more complex, globally distributed and far more interconnected than public and academic discourses allow us to perceive. This necessitates a far broader, more multi-, and interdisciplinary view than any single academic field or practice can offer. To tackle all the known and yet to be discovered problems related to AI—whether cultural, technological, societal, or environmental—a view that can account for more complexity is needed. I started this by offering visualisations of the various sociomaterial and semiotic relationships that AI and computer systems generally have. We saw how the Expected AI is still largely constructed and informed by science fiction and various cultural code-texts.

To understand ourselves and the world around us, we do not need AI to "simplify" it for us. Instead, we need to learn better ways to handle the complexity of the world—by ourselves—and fast. Semiotics can help, starting with the perspective offered by my Tartu semiotics colleague Merit Rickberg who defended her PhD thesis just a few months earlier. Building extensively on the theory of Juri Lotman, she proposes a new strand of complexity thinking "to serve as a platform for translation between different areas of knowledge" (Rickberg 2023: 14). Following Kobus Marais, Rickberg argues that "the complexity perspective is inherently ecological in that it sees the whole of reality as interrelated, having emerged from the physical" (Rickberg 2023: 14). As I argued in chapter 2.4, such a view is direly needed to conceptualise and understand the global material roots and impact of any AI system. At the same time, complexity thinking is already built on several other research traditions, including AI, but also cybernetics, and various strands of systems theory (Rickberg 2023: 18). In this context, it may be easy to assume that the field of AI already includes these tools to address its impending issues. However, as Philip Agre (1995, 1997a) thoroughly demonstrates, AI as a field still lacks necessary meta-perspectives and tools to distinguish between different levels of inquiry. AI also lacks affinities to humanities and social sciences—a schism that in this century and decade only seems to become more pronounced. Therefore, a new, self-reflective, and self-critical meta-level perspective such as semiotics-inspired complexity thinking is necessary, especially given that AI systems have arguably huge, disproportionate, and largely unstudied impact on global climate. Therefore, as a next step, I intend to develop the Ecologies of AI model further, including many specific case studies and interdisciplinary teams to address various problems visualised in the model.

Aimee van Wynsberghe's (2021) approach to Sustainable AI could partly provide a framework to assess and challenge the product chain of AI with its environmental and social impact. In making sense of the globality of AI, Immanuel Wallerstein's *world system theory* (Wallerstein 2004) combined with Juri Lotman's *semiosphere* with its similar core–periphery dynamics (Lotman 2005, 2009) offer useful avenues for further exploration. Walter Mignolo's perspectives on geopolitics of knowledge and decoloniality (Mignolo 2002, 2011, 2021; Mignolo and Walsh 2018) provide paths for seeking more just reframing instead of the current oppressive objectivations of AI.

The public discourse on AI has significantly heated up in recent years, amounting to "Twitter wars" waged in social media largely between AI researchers and social scientists. Indeed, as Agre argued in 1997, AI makers are usually unable to take any critique whatsoever unless it begins with a full-scale appreciation of all the hard and ingenious work they have done. Only in previous decades, their precious inventions were confined to their laboratories, only reaching the public through discursive practices. Now, they are regularly being unleashed upon the world—often in their testing phase—and social scientists are angry because they can regularly observe how these unregulated technologies and platforms interfere with various cultures, destabilize social institutions, and destroy the lives of the less privileged to an extent that sometimes amounts to various catastrophes on a global scale. 2023 will go down in history as the year of the corporate deep-learning and chatbots "arms race"[43]. It will also be the moment when the Internet transitions from being a collection of largely useful cultural resources to pulling all the connected AI systems into a race toward model collapse. Who else goes down with this avalanche, we do not yet know.

In addition to all the problems with AI technologies, their reception and interpretations, the discourse—academic and public—on technology and AI has also led to several significant contributions in multiple research fields precisely due to its liminality and controversiality. The positioning of AI as human's Other, and thus inviting the conflation of machine and human ontologies has helped demonstrate the limitations of many scientific theories and challenged several leading paradigms (behaviourism, computationalism, etc) to rethink their models.

Understanding history as old as Ancient Greece in the context of AI is also still important. We saw that historians of AI like to draw parallels back to Antikythera and Greek legends (Liveley and Thomas 2020, see also Mayor 2018), as if to say that "thinking machines" have been a most natural ambition of human culture since times immemorial. While I agree with Tatiana Bur in that the technical schemata or the cultural functions of Heron's temple gates and today's neural networks could not be more different, there is something in the ancient power-and-politics dynamic that we cannot seem to shake completely either. The

[43] This is a metaphor of how the problem is framed in discourse, and not a suggestion to compare the situation to a real arms race; but given the frenzy of launching new language models every few weeks or months, the metaphor may well ring true—if words can be considered weapons. Some argue that they can.

reception of certain AI products in today's culture and media space seems to reflect the logic of *thauma idesthai* from Ancient Greek religion: it is made clear that the AI marvels we observe result from the ingenuity of human engineering alone, and yet we cannot help but treat them as if there was a divine presence within. How else to explain the widespread anthropomorphism of robotics and language models? Certainly, it can be a variety of "mock" anthropomorphism, simply "because we find it easier to deal with machines in human terms [and] not because machines share our experiences" (de Waal 1999: 270). In this context, however, it is still not entirely reasonable and understandable why, for example, as I analyse in Papers II and III, media outlets keep quoting a computer-generated text as if it were the "opinion" of "GPT-3 the robot"[44], or people keep persisting in trying to "have a conversation with the bot"[45].

---

[44] At the time of writing (May 2023), more recent examples of such discourse include the reception of ChatGPT, Bing, and other search-and-chat engines.

[45] Karen Weise, "Microsoft's Bing Chatbot Offers Some Puzzling and Inaccurate Responses." *The New York Times*, February 15, 2023, https://www.nytimes.com/2023/02/15/technology/microsoft-bing-chatbot-problems.html. Accessed October 15, 2023; Kevin Roose, "A Conversation With Bing's Chatbot Left Me Deeply Unsettled." *The New York Times*, February 16, 2023, https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html. Accessed October 15, 2023; Kevin Roose, "Bing's A.I. Chat: 'I Want to Be Alive. 😈.'" *The New York Times*, February 16, 2023, https://www.nytimes.com/2023/02/16/technology/bing-chatbot-transcript.html. Accessed October 15, 2023.

Lubuya looked at me as if I were a fool. "Every day people are dying because of the cobalt. Describing this will not change anything."

—Siddharth Kara, Cobalt Red

REFERENCES

Adam, Alison 1998. *Artificial Knowing: Gender and the Thinking Machine*. London; New York: Routledge.

Agre, Philip E. 1995. The Soul Gained and Lost: Artificial Intelligence as a Philosophical Project. *Stanford Humanities Review* 4(2): 1–19.

——— 1997a. Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI. In: Bowker, G.; Star, S. L.; Gasser, L.; Turner, W. (eds.), *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide*. New York, London: Psychology Press, 131–157.

——— 1997b. *Computation and Human Experience*. Cambridge; New York: Cambridge University Press.

——— 2002. Real-Time Politics: The Internet and the Political Process. *The Information Society* 18(5): 311–331.

Ali, Syed Mustafa 2017. Transhumanism and/as Whiteness. In: *Proceedings of the IS4SI 2017 Summit DIGITALISATION FOR A SUSTAINABLE SOCIETY, Gothenburg, Sweden, 12–16 June 2017.*

——— 2019. "White Crisis" and/as "Existential Risk," or the Entangled Apocalypticism of Artificial Intelligence. *Zygon*® 54(1): 207–224.

Ambrosetti, Nadia 2010. *Cultural Roots of Technology: An Interdisciplinary Study of Automated Systems from the Antiquity to the Renaissance*. Milano: Università degli Studi di Milano.

Andersen, Peter Bøgh, Per Hasle, and Per A. Brandt 1997. Machine Semiosis. In Posner, R.; Robering, K.; Sebeok, T. A. (eds.), *Semiotik: Ein Handbuch Zu Den Zeichentheoretischen Grundlagen von Natur Und Kultur = Semiotics: A Handbook on the Sign-Theoretic Foundations of Nature and Culture*. Berlin; New York: Walter de Gruyter, 548–571.

Andrae, Anders S. G., and Tomas Edler 2015. On Global Electricity Usage of Communication Technology: Trends to 2030. *Challenges* 6(1): 117–157.

Appio, Francesco, Davide La Torre, Francesca Lazzeri, Francesco Schiavone, and Hatem Masri 2023. *Impact of Artificial Intelligence in Business and Society: Opportunities and Challenges*. Abingdon, Oxon; New York, NY: Routledge.

Araujo, Theo, Natali Helberger, Sanne Kruikemeier, and Claes H. de Vreese 2020. In AI We Trust? Perceptions about Automated Decision-Making by Artificial Intelligence. *AI & SOCIETY* 35(3): 611–623.

Ashby, W. Ross 1956. *An Introduction to Cybernetics*. London: Chapman & Hall.

Atkinson, Robert D, Doug Brake, Daniel Castro, Colin Cunliff, Joe Kennedy, Michael McLaughlin, Alan McQuinn, and Joshua New 2019. *A Policymaker's Guide to the "Techlash"—What It Is and Why It's a Threat to Growth and Progress*. Information Technology & Innovation Foundation.

Austin, J. L. 1962. How to Do Things with Words: The William James Lectures Delivered in Harvard University in 1955. Oxford: Oxford University Press.

Belkhir, Lotfi, and Ahmed Elmeligi 2018. Assessing ICT Global Emissions Footprint: Trends to 2040 & Recommendations. *Journal of Cleaner Production* 177: 448–463.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. Virtual Event Canada: ACM, 610–623.

Benjamin, Ruha 2013. *People's Science: Bodies and Rights on the Stem Cell Frontier*. Stanford, California: Stanford University Press.
——— 2019a. *Race after Technology: Abolitionist Tools for the New Jim Code*. Medford, MA: Polity.
——— (ed.) 2019b. *Captivating Technology: Race, Carceral Technoscience, and Liberatory Imagination in Everyday Life*. Durham: Duke University Press.
Bennett, Tyler James 2021. *Detotalization and Retroactivity: Black Pyramid Semiotics*. Dissertationes semioticae Universitatis Tartuensis, 35. University of Tartu.
Berk, Richard A. 2021. Artificial Intelligence, Predictive Policing, and Risk Assessment for Law Enforcement. *Annual Review of Criminology* 4(1): 209–237.
Bieser, Jan C. T., Ralph Hintemann, Lorenz M. Hilty, and Severin Beucker 2023. A Review of Assessments of the Greenhouse Gas Footprint and Abatement Potential of Information and Communication Technology. *Environmental Impact Assessment Review* 99: 107033.
Bihouix, Philippe 2014. *L'âge des low tech: vers une civilisation techniquement soutenable*. Paris: Éditions du Seuil.
——— 2020. *The Age of Low Tech: Towards a Technologically Sustainable Civilization*. (McMahon, C., transl.) Bristol: Bristol University Press.
Bijker, Wiebe E. 1995. *Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change*. Cambridge, Mass: MIT Press.
Birhane, Abeba 2021. Algorithmic Injustice: A Relational Ethics Approach. *Patterns* 2(2): 100205.
Blackwell, Alan F. 2021. Ethnographic Artificial Intelligence. *Interdisciplinary Science Reviews* 46(1–2): 198–211.
Born, Rainer, and Ilse Born-Lechleitner 1987. Introduction. In: Born, R. (ed.), *Artificial Intelligence: The Case Against*. New York: St. Martin's Press, i–xxxv.
Bremmer, Jan N. 1994. *Greek Religion*. Oxford; New York: Oxford University Press.
Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei 2020. Language Models Are Few-Shot Learners. *arXiv:2005.14165 [cs]*.
Bryson, Joanna J. 2010. Robots Should Be Slaves. In: Wilks, Yorik (ed.), *Close Engagements with Artificial Companions: Key social, psychological, ethical and design issues.* Amsterdam: Benjamins, 63–74.
Bryson, Joanna J., and Philip P. Kime 2011. Just an Artifact: Why Machines Are Perceived as Moral Agents. In: *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence – Volume Two*. Barcelona, Catalonia, Spain: AAAI Press, 1641–1646.
Buolamwini, Joy, and Timnit Gebru 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research* 81: 1–15.
Bur, Tatiana 2016. *Mechanical Miracles: Automata in Ancient Greek Religion*. Master's Thesis. University of Sydney.
——— 2020. *Manufacturing the Marvellous: Technology and Divine Presence in Ancient Greek Religion*. Doctoral thesis. University of Cambridge.

Burghardt, Gordon M. 1991. Cognitive Ethology and Critical Anthropomorphism: A Snake with Two Heads and Hognose Snakes That Play Dead. In: Ristau, C. A. (ed.), *Cognitive Ethology: The Minds of Other Animals: Essays in Honor of Donald R. Griffin*. Hillsdale, New Jersey; Hove and London: Lawrence Erlbaum, 53–90.

Cadena, Marisol de la, and Mario Blaser (eds.) 2018. *A World of Many Worlds*. Durham: Duke University Press.

Cave, Stephen, Claire Craig, Kanta Sarasvati Dihal, Sarah Dillon, Jessica Montgomery, Beth Singler, and Lindsay Taylor 2018. *Portrayals and Perceptions of AI and Why They Matter*. London: The Royal Society.

Cave, Stephen, and Kanta Dihal 2018. Ancient Dreams of Intelligent Machines: 3,000 Years of Robots. *Nature* 559(7715): 473–475.

——— 2019. Hopes and Fears for Intelligent Machines in Fiction and Reality. *Nature Machine Intelligence* 1(2): 74–78.

——— 2020. The Whiteness of AI. *Philosophy & Technology* 33(4): 685–703.

Cave, Stephen, Kanta Sarasvati Monique Dihal, and Sarah Dillon (eds.) 2020. *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*. Oxford: Oxford University Press.

Chandler, Daniel 2017. *Semiotics: The Basics*. Abingdon Oxon; New York, NY: Routledge.

Chesterman, Simon 2021. *We, the Robots? Regulating Artificial Intelligence and the Limits of the Law*. Cambridge, UK; New York, USA: Cambridge University Press.

Chu, Seo-Young 2010. *Do Metaphors Dream of Literal Sleep?: A Science-Fictional Theory of Representation*. Cambridge, MA; London: Harvard University Press.

Cicero, Marcus Tullius 1967. *De Natura Deorum*. (Rackham, H., trans.) Cambridge, Mass: Harvard University Press.

Colapietro, Vincent, Winfried Nöth, and Guilherme Henrique de Oliveira Cestari 2020. A Dialogue on Cognitive Semiotics: Minds, and Machines. *TECCOGS: Revista Digital de Tecnologias Cognitivas* (21).

Cooke, Conrad William 1893. *Automata Old and New*. London: Chiswick Press.

Couldry, Nick, and Andreas Hepp 2017. *The Mediated Construction of Reality*. Cambridge, UK; Malden, MA: Polity Press.

Crawford, Kate 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press.

Csicsery-Ronay, Istvan 2004. Who Framed Science Fiction? *Science Fiction Studies* 31(1): 122–126.

Danowski, Déborah, and Eduardo Viveiros de Castro 2017. *The Ends of the World*. Malden, MA: Polity.

Dhar, Payal 2020. The Carbon Impact of Artificial Intelligence. *Nature Machine Intelligence* 2(8): 423–425.

Dinello, Daniel 2005. *Technophobia! Science Fiction Visions of Posthuman Technology*. Austin: University of Texas Press.

Dreher, Jochen 2016. The Social Construction of Power: Reflections Beyond Berger/Luckmann and Bourdieu. *Cultural Sociology* 10(1): 53–68.

Dydynski, Jason Mario 2021. *Semiotic Modeling of Cuteness in Cartoon Characters/Mascots*. Dissertationes semioticae Universitatis Tartuensis, 36. University of Tartu.

Dydynski, Jason Mario, and Nelly Mäekivi 2021. Impacts of Cartoon Animals on Human–Alloanimal Relations. *Anthrozoös* 34(6): 753–766.

Eco, Umberto 1976. *A Theory of Semiotics*. Bloomington; London: Indiana University Press.

——— 1979. *The Role of the Reader: Explorations in the Semiotics of Texts*. Bloomington: Indiana University Press.

Ekbia, Hamid, and Bonnie Nardi 2016. Social Inequality and HCI: The View from Political Economy. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems,* 4997–5002.

Elias, Norbert 1978. *What Is Sociology?* London: Hutchinson.

Emmeche, Claus 2007. A Biosemiotic Note on Organisms, Animals, Machines, Cyborgs, and the Quasi-Autonomy of Robots. *Pragmatics & Cognition* 15(3): 455–483.

Epley, Nicholas, Adam Waytz, and John T. Cacioppo 2007. On Seeing Human: A Three-Factor Theory of Anthropomorphism. *Psychological Review* 114(4): 864–886.

Eubanks, Virginia 2017. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.

Falk, Sophia, and Aimee van Wynsberghe 2023. Challenging AI for Sustainability: What Ought It Mean? *AI Ethics*. doi: 10.1007/s43681-023-00323-3

Fast, Ethan, and Eric Horvitz 2017. Long-Term Trends in the Public Perception of Artificial Intelligence. In: *Proceedings of the AAAI Conference on Artificial Intelligence*.

Favareau, Donald (ed.) 2010. *Essential Readings in Biosemiotics: Anthology and Commentary*. Dordrecht; New York: Springer.

——— 2015. Symbols Are Grounded Not in Things, but in Scaffolded Relations and Their Semiotic Constraints (Or How the Referential Generality of Symbol Scaffolding Grows Minds). *Biosemiotics* 8(2): 235–255.

Floridi, Luciano 2020. AI and Its New Winter: From Myths to Realities. *Philosophy & Technology* 33(1): 1–3.

——— 2023. On Good and Evil, the Mistaken Idea That Technology Is Ever Neutral, and the Importance of the Double-Charge Thesis. SSRN Scholarly Paper. doi: 10.2139/ssrn.4551487

Freitag, Charlotte, Mike Berners-Lee, Kelly Widdicks, Bran Knowles, Gordon S. Blair, and Adrian Friday 2021. The Real Climate and Transformative Impact of ICT: A Critique of Estimates, Trends, and Regulations. *Patterns* 2(9): 100340.

Gebru, Timnit, and Émile P. Torres (forthcoming). Eugenics and the Promise of Utopia through Artificial General Intelligence. *First Monday*.

Gell, Alfred 1992. The Technology of Enchantment and the Enchantment of Technology. In: Coote, Jeremy; Sheldon, Anthony (eds.), *Anthropology, Art, and Aesthetics*. Oxford: Oxford University Press, 40–63.

Geraci, Robert M. 2008. Apocalyptic AI: Religion and the Promise of Artificial Intelligence. *Journal of the American Academy of Religion* 76(1): 138–166.

Gerolemou, Maria 2019. Technological Wonder in Herodotus' Histories. In: Kazantzidis, G. (ed.), *Medicine and Paradoxography in the Ancient World*. Berlin; Boston: De Gruyter, 41–52.

Gherlone, Laura 2023. On Binary Opposition and Binarism: A Long-Distance Dialogue between Decolonial Critique and the Lotmanian Semiotics. *Sign Systems Studies* 51(2): 254–279.

Glickman, Moshe, and Tali Sharot 2022. Biased AI Systems Produce Biased Humans. OSF Preprints. https://osf.io/c4e7r/. Accessed on October 23, 2023.

Goddard, Kate, Abdul Roudsari, and Jeremy C Wyatt 2012. Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators. *Journal of the American Medical Informatics Association* 19(1): 121–127.

Gómez-Barris, Macarena 2017. *The Extractive Zone: Social Ecologies and Decolonial Perspectives*. Durham; London: Duke University Press. https://www2.datainnovation.org/2023-ai-act-technology-neutral.pdf. Accessed on October 23, 2023.
Grady, Patrick 2023. *The AI Act Should Be Technology-Neutral*. Center for Data Innovation.
Gray, Chris Hables 1997. Artificial Intelligence at War: An Analysis of the Aegis System in Combat. In: Agre, P. E.; Schuler, D. (eds.), *Reinventing Technology, Rediscovering Community: Critical Explorations of Computing as a Social Practice*. London: Ablex Publishing Corporation, 127–142.
Grillo, Francesco 2019. *Hero of Alexandria's Automata: A Critical Edition and Translation, Including a Commentary on Book One*. PhD Thesis. University of Glasgow.
Grünke, Paul 2020. Chess, Artificial Intelligence, and Epistemic Opacity. *Információs Társadalom* 19(4): 7–17.
Habgood-Coote, Joshua 2023. Deepfakes and the Epistemic Apocalypse. *Synthese* 201(3): 103.
Hajian, Sara, Francesco Bonchi, and Carlos Castillo 2016. Algorithmic Bias: From Discrimination Discovery to Fairness-Aware Data Mining. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery, 2125–2126.
Hansen, Lene, and Helen Nissenbaum 2009. Digital Disaster, Cyber Security, and the Copenhagen School. *International Studies Quarterly* 53(4): 1155–1175.
Harbers, Maaike, Marieke M. M. Peeters, and Mark A. Neerincx 2017. Perceived Autonomy of Robots: Effects of Appearance and Context. In Aldinhas Ferreira, M. J.; Silva Sequeira, J.; Tokhi, M. O.; Kadar, E. E.; Virk, G. S. (eds.), *A World with Robots. International Conference on Robot Ethics: ICRE 2015.* Intelligent Systems, Control and Automation: Science and Engineering 84. Cham: Springer International Publishing, 19–33.
Hare, Stephanie 2022. *Technology Is Not Neutral: A Short Guide to Technology Ethics*. London: London Publishing Partnership.
Hassine, Tsila, and Ziv Neeman 2019. The Zombification of Art History: How AI Resurrects Dead Masters, and Perpetuates Historical Biases. *Journal of Science and Technology of the Arts* 11(2): 28–35.
Héder, Mihály 2023. The Epistemic Opacity of Autonomous Systems and the Ethical Consequences. *AI & SOCIETY* 38(5): 1819–1827.
Hicks, Mar 2017. *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing*. Cambridge, MA: MIT Press.
Hornborg, Alf 2021. Machines as Manifestations of Global Systems: Steps toward a Sociometabolic Ontology of Technology. *Anthropological Theory* 21(2): 206–227.
Horowitz, Michael C. 2019. When Speed Kills: Lethal Autonomous Weapon Systems, Deterrence and Stability. *Journal of Strategic Studies* 42(6): 764–788.
Horowitz, Michael C., and Lauren Kahn 2023. Bending the Automation Bias Curve: A Study of Human and AI-Based Decision Making in National Security Contexts. arXiv:2306.16507
Hussen, Ahmed M. 2004. *Principles of Environmental Economics*. London: Routledge.
Irvine, Martin 2022. Semiotics in Computing and Information Systems. In Pelkey, J.; Walsh Matthews, S. (eds.), *Bloomsbury Semiotics. Volume 2: Semiotics in the Natural and Technical Sciences*. London: Bloomsbury Academic, 203–37.
Jazari, Isma`il ibn al-Razzaz 1974. *The Book of Knowledge of Ingenious Mechanical Devices*. (Hill, D. R., trans.) Dordrecht; Boston: D. Reidel Publishing Company.

Joerges, Bernward 1999. Do Politics Have Artefacts? *Social Studies of Science* 29(3): 411–31.

Johansen, Jørgen Dines 2002. *Literary Discourse: A Semiotic-Pragmatic Approach to Literature*. Toronto; Buffalo: University of Toronto Press.

Johansen, Sigrid Redse 2018. So Man Created Robot in His Own Image: The Anthropomorphism of Autonomous Weapon Systems and the Law of Armed Conflict. *Oslo Law Review* 5(2): 89–102.

Johnson, Deborah G., and Mario Verdicchio 2017. Reframing AI Discourse. *Minds and Machines* 27(4): 575–590.

——— 2018. Why Robots Should Not Be Treated like Animals. *Ethics and Information Technology* 20(4): 291–301.

Johnson, James 2022. Delegating Strategic Decision-Making to Machines: Dr. Strangelove Redux? *Journal of Strategic Studies* 45(3): 439–477.

Johnson, Jim 1988. Mixing Humans and Nonhumans Together: The Sociology of a Door-Closer*. *Social Problems* 35(3): 298–310.

Johnson, Thaddeus L., Natasha N. Johnson, Denise McCurdy, and Michael S. Olajide 2022. Facial Recognition Systems in Policing and Racial Disparities in Arrests. *Government Information Quarterly* 39(4): 101753. doi:10.1016/j.giq.2022.101753

Jorna, René J., Barend van Heusden, and Roland Posner (eds.) 1993. *Signs, Search and Communication: Semiotic Aspects of Artificial Intelligence*. Berlin; New York: Walter de Gruyter.

Kahyaoğlu, Sezer Bozkuş (ed.) 2021. *The Impact of Artificial Intelligence on Governance, Economics and Finance, Volume I*. Singapore: Springer Nature Singapore.

Kapania, Shivani, Oliver Siy, Gabe Clapper, Azhagu SP, and Nithya Sambasivan 2022. "Because AI Is 100% Right and Safe": User Attitudes and Sources of AI Authority in India. In: *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 18 pages. doi: 10.1145/3491102.3517533

Kara, Siddharth 2014. *Bonded Labor: Tackling the System of Slavery in South Asia*. New York: Columbia University Press.

——— 2017. *Modern Slavery: A Global Perspective*. New York: Columbia University Press.

——— 2023. *Cobalt Red: How the Blood of the Congo Powers Our Lives*. New York: St. Martin's Press.

Katz, Yarden 2020. *Artificial Whiteness: Politics and Ideology in Artificial Intelligence*. New York: Columbia University Press.

Kennedy, J. S. 1992. *The New Anthropomorphism*. Cambridge [England]; New York: Cambridge University Press.

Keyes, Os 2020. Automating Autism: Disability, Discourse, and Artificial Intelligence. *The Journal of Sociotechnical Critique* 1(1): 1–31. doi:10.25779/89bj-j396

Kranzberg, Melvin 1986. Technology and History: "Kranzberg's Laws." *Technology and Culture* 27(3): 544–560.

Laan, Andres 2017. *Tehisintellekt: loomadest ja masinatest*. Tallinn: Koolibri.

Laclau, Ernesto 2005. *On Populist Reason*. London; New York: Verso.

Lacoste, Alexandre, Alexandra Luccioni, Victor Schmidt, and Thomas Dandres 2019. Quantifying the Carbon Emissions of Machine Learning. arXiv:1910.09700.

Latour, Bruno, and Steve Woolgar 1986. *Laboratory Life: The Construction of Scientific Facts*. Princeton, NJ: Princeton University Press.

Le Guin, Ursula K. 2010. *The Left Hand of Darkness*. New York: Ace Books.

Lee, Jyh-An, Reto M. Hilty, and Kung-Chung Liu (eds.) 2021. *Artificial Intelligence and Intellectual Property*. Oxford; New York: Oxford University Press.
Leveson, Nancy G. 2017. The Therac-25: 30 Years Later. *Computer* 50(11): 8–11.
Lévi-Strauss, Claude 1950. L'øeuvre de Marcel Mauss. *Cahiers internationaux de sociologie* 8: 72–112.
——— 1987. *Introduction to the Work of Marcel Mauss*. London: Routledge & Kegan Paul.
Li, Pengfei, Jianyi Yang, Mohammad A. Islam, and Shaolei Ren 2023. Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models. arXiv:2304.03271.
Ligozat, Anne-Laure, Julien Lefevre, Aurélie Bugeau, and Jacques Combaz 2022. Unraveling the Hidden Environmental Impacts of AI Solutions for Environment Life Cycle Assessment of AI Solutions. *Sustainability* 14(9): 5172.
Liveley, Genevieve, and Sam Thomas 2020. Homer's Intelligent Machines: AI in Antiquity. In: Cave, S.; Dihal, K. S. M.; Dillon, S. (eds.), *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*. Oxford: Oxford University Press, 25–48.
Lotman, Juri 2005. On the Semiosphere. *Sign Systems Studies* 33(1): 205–229.
——— 2009. *Culture and Explosion*. Berlin; New York: Mouton de Gruyter.
Luccioni, Sasha, Sylvain Viguier, and Anne-Laure Ligozat 2022. Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model. arXiv:2211.02001v1
Madisson, Mari-Liis, and Andreas Ventsel 2022. Lotman and Social Media Studies. In: Tamm, Marek; Torop, Peeter (eds.), *The Companion to Juri Lotman: A Semiotic Theory of Culture*. London; New York: Bloomsbury Academic, 446–460.
Marais, Kobus 2019. "Effects Causing Effects": Considering Constraints in Translation. In: Marais, K.; Meylaerts (eds.), *Complexity Thinking in Translation Studies: Methodological Considerations*, edited by K. Marais and R. Meylaerts. New York, NY: Routledge, 53–72.
Masís, Jethro 2014. Making AI Philosophical Again: On Philip E. Agre's Legacy. *Continent* 4(1): 58–70.
Mayor, Adrienne 2018. *Gods and Robots: Myths, Machines, and Ancient Dreams of Technology*. Princeton: Princeton University Press.
Mays, Wolfe 1952. Can Machines Think? *Philosophy* 27(101): 148–162.
——— 2001. My Reply to Turing: Fiftieth Anniversary. *Journal of the British Society for Phenomenology* 32(1): 4–23.
Mbembe, Achille 2019. *Necropolitics*. (Corcoran, S., trans.) Durham London: Duke University Press.
McCarthy, John 1988. Bloomfield, Brian, ed. The Question of Artificial Intelligence: Philosophical and Sociological Perspectives (Book Review). *Annals of the History of Computing* 10(3): 224–229.
McCarthy, John, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon 1955. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf. Accessed on September 18, 2023.
——— 2006. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI magazine* 27(4): 12–14.
McCorduck, Pamela 2004. *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*. Natick, Mass: A.K. Peters.

McMillan-Major, Angelina, Emily M. Bender, and Batya Friedman 2023. Data Statements: From Technical Concept to Community Practice. *ACM Journal on Responsible Computing*. doi:10.1145/3594737
Meadows, Donella H., Dennis L. Meadows, Jørgen Randers, and William W. Behrens 1972. *The Limits to Growth: A Report for the Club of Rome's Project on the Predicament of Mankind*. New York: Universe Books.
Menise, Tatjana 2019. Fairy Tales between Transformation and Repetition: How Audiences Rethink the Big Romantic Myth through Disney Princess Stories. *Sign Systems Studies* 47(3/4): 526–551.
Meunier, Jean-Guy 1989. Artificial Intelligence and Sign Theory. *Semiotica* 77(1–3): 43–64.
——— 2021. *Computational Semiotics*. New York: Bloomsbury Academic.
Mignolo, Walter D. 2002. The Geopolitics of Knowledge and the Colonial Difference. *The South Atlantic Quarterly* 101(1): 57–96.
——— 2011. *The Darker Side of Western Modernity: Global Futures, Decolonial Options*. Durham: Duke University Press.
——— 2021. *The Politics of Decolonial Investigations*. Durham: Duke University Press.
Mignolo, Walter D., and Arturo Escobar 2010. *Globalization and the Decolonial Option*. London New York: Routledge.
Mignolo, Walter D., and Catherine Walsh 2018. *On Decoloniality: Concepts, Analytics, Praxis*. Durham NC; London: Duke University Press.
Mills, Charles W. 1997. *The Racial Contract*. Ithaca (N.Y.): Cornell University Press.
More, Trenchard 1956a. Computer Decisions in Deductive Logic. https://raysolomonoff.com/dartmouth/boxa/moreprop.pdf. Accessed on September 18, 2023.
——— 1956b. Dartmouth Conference on Artificial Intelligence. Report on the 5th and 6th Weeks. http://raysolomonoff.com/dartmouth/boxa/dart56more5th6thweeks.pdf. Accessed on September 18, 2023.
——— 1973. Axioms and Theorems for a Theory of Arrays. *IBM Journal of Research and Development* 17(2): 135–175.
Mosco, Vincent 2004. *The Digital Sublime: Myth, Power, and Cyberspace*. Cambridge, Mass: MIT Press.
Mosier, Kathleen L., and Linda J. Skitka 1999. Automation Use and Automation Bias. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 43(3): 344–348.
Musa, Roberto 2019. Myths, Tests and Games: Cultural Roots and Current Routes of Artificial Intelligence. Doctorado en Psicología. Pontificia Universidad Católica de Chile.
——— 2020. Echoes of Myth and Magic in the Language of Artificial Intelligence. *AI & SOCIETY* 35: 1009–1024.
Nadarajan, Gunalan 2008[2007]. *Islamic Automation: A Reading of al-Jazari's The Book of Knowledge of Ingenious Mechanical Devices (1206)*. Foundation for Science Technology and Civilisation. https://muslimheritage.com/islamic-automation-al-jazaris-book-of-knowledge-of-ingenious-mechanical-devices/. Accessed on September 18, 2023.
Nadin, Mihai 2007. Semiotic Machine. *Public Journal of Semiotics* 1(1): 57–75.
——— 2011. Information and Semiotic Processes the Semiotics of Computation. *Cybernetics & Human Knowing* 18(1–2): 153–175.
——— 2019. Machine Intelligence: A Chimera. *AI & SOCIETY* 34(2): 215–242.

Nat 2023. AI's Carbon Footprint: Understanding and Reducing the Environmental Impact of Large Models. The AI Observer. Substack newsletter. https://theaiobserverx.substack.com/p/ais-carbon-footprint-understanding. Accessed on September 18, 2023.
Natale, Simone, and Andrea Ballatore 2020. Imagining the Thinking Machine: Technological Myths and the Rise of Artificial Intelligence. *Convergence* 26(1): 3–18.
Naudé, Wim 2021. Artificial Intelligence: Neither Utopian nor Apocalyptic Impacts Soon. *Economics of Innovation and New Technology* 30(1): 1–23.
Nauta, Doede 1972. *The Meaning of Information*. The Hague; Paris: Mouton.
Ndlovu-Gatsheni, Sabelo J. 2018. *Epistemic Freedom in Africa: Deprovincialization and Decolonization*. London; New York: Routledge.
Nilsson, Nils J. 2009. *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. Cambridge: Cambridge University Press.
Noble, David F. 1997. *The Religion of Technology: The Divinity of Man and the Spirit of Invention*. New York: A.A. Knopf.
Nöth, Winfried 2001. Semiosis and the Umwelt of a Robot. *Semiotica* 134(1/4): 695–699.
——— 2002. Semiotic Machines. *Cybernetics & Human Knowing* 9(1): 5–21.
——— 2008. Sign Machines in the Framework of Semiotics Unbounded. *Semiotica* 2008(169): 319–341.
——— 2019. Peirce on the Continuity between Human and Nonhuman Minds. In: Middelhoff, F. (ed.), *Texts, Animals, Environments: Zoopoetics and Ecopoetics*. Freiburg: Rombach Verlag, 139–155.
Oberdiek, Hans 1990. Technology: Autonomous or Neutral. *International Studies in the Philosophy of Science* 4(1): 67–77.
O'Neil, Cathy 2017. *Weapons of Math Destruction: How Big Data Produces Inequality and Threatens Democracy*. London: Penguin Books.
OpenAI 2023. *GPT-4 Technical Report*. OpenAI. https://cdn.openai.com/papers/gpt-4.pdf. Accessed on September 18, 2023.
Pamuk, Zeynep 2021. Risk and Fear: Restricting Science under Uncertainty. *Journal of Applied Philosophy* 38(3): 444–460.
Parasuraman, Raja, and Dietrich H. Manzey 2010. Complacency and Bias in Human Use of Automation: An Attentional Integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 52(3): 381–410.
Payne, Kenneth 2021. *I, Warbot: The Dawn of Artificially Intelligent Conflict*. London: Hurst & Company.
Peirce, Charles Sanders 1887. Logical Machines. *The American Journal of Psychology* 1(1): 165–170.
Pickover, Clifford A. 2019. *Artificial Intelligence: An Illustrated History. From Medieval Robots to Neural Networks*. New York: Sterling.
Pinsky, Leonard 1951. Do Machines Think about Machines Thinking? *Mind* LX: 397–398.
Pirenne-Delforge, Vinciane 2010. Reading Pausanias: Cults of the Gods and Representation of the Divine. In: Bremmer, Jan Nicholas; Erskine, Andrew (eds.), *The Gods of Ancient Greece: Identities and Transformations*. Edinburgh: Edinburgh University Press, 375–387.
Pitron, Guillaume 2023. *The Rare Metals War: The Dark Side of Clean Energy and Digital Technologies*. (Jacobsohn, B., trans.) London: Scribe Publications.
Polanyi, Michael 1958. *Personal Knowledge: Towards a Post-Critical Philosophy*. Chicago: University of Chicago Press.
——— 1966. *The Tacit Dimension*. Garden City, N.Y: Doubleday.

Posner, Roland, Klaus Robering, and Thomas A. Sebeok (eds.) 1997. *Semiotik: Ein Handbuch Zu Den Zeichentheoretischen Grundlagen von Natur Und Kultur = Semiotics: A Handbook on the Sign-Theoretic Foundations of Nature and Culture. Vol.1.* Berlin; New York: Walter de Gruyter.
——— (eds.) 1998. *Semiotik: Ein Handbuch Zu Den Zeichentheoretischen Grundlagen von Natur Und Kultur = Semiotics: A Handbook on the Sign-Theoretic Foundations of Nature and Culture. Vol. 2*. Berlin; New York: Walter de Gruyter.
——— (eds.) 2003. *Semiotik: ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur = Semiotics: a handbook on the sign-theoretic foundations of nature and culture. Vol. 3*. Berlin: de Gruyter.
——— (eds.) 2004. *Semiotik: Ein Handbuch Zu Den Zeichentheoretischen Grundlagen von Natur Und Kultur = Semiotics: A Handbook on the Sign-Theoretic Foundations of Nature and Culture. Vol. 4*. Berlin, New York: Walter de Gruyter.
Prem, Erich 1998. Semiosis in Embodied Autonomous Systems. In: *Proceedings of the 1998 IEEE International Symposium on Intelligent Control (ISIC) Held Jointly with IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA) Intell*, 724–729.
Propp, Vladimir 1968. *Morphology of the Folktale*. Austin: University of Texas Press.
——— 1984. *Theory and History of Folklore*. Minneapolis: University of Minnesota Press.
——— 2012. *The Russian Folktale by Vladimir Yakovlevich Propp*. (Forrester, S. E. S., trans.) Detroit: Wayne State University Press.
——— 2022. *Imemuinasjutu ajaloolised juured*. (Arukask, Madis, trans.) Tartu: Ilmamaa.
Prou, Victor 1884. Les théâtres d'automates en Grèce au IIe siècle avant l'ère chrétienne d'après les Αὐτοματοποιῖκὰ d'Héron d'Alexandrie. *Mémoires présentés par divers savants étrangers à l'Académie* 9(2): 117–274.
Puumeister, Ott, and Silver Rattasepp 2022. Lõpuaja lõputus. *Akadeemia* 9: 1539–1555.
Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever 2019. Language Models Are Unsupervised Multitask Learners. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf. Accessed on September 18, 2023.
Raji, Inioluwa Deborah, Timnit Gebru, Margaret Mitchell, Joy Buolamwini, Joonseok Lee, and Emily Denton 2020. Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. New York: ACM, 145–151.
Raji, Inioluwa Deborah, I. Elizabeth Kumar, Aaron Horowitz, and Andrew Selbst 2022. The Fallacy of AI Functionality. In: *2022 ACM Conference on Fairness, Accountability, and Transparency*. Seoul, Republic of Korea: Association for Computing Machinery, 959–972.
Richards, Neil M., and William D. Smart 2016. How Should the Law Think about Robots? In: Calo, Ryan; Froomkin, A. Michael; Kerr, Ian (eds.), *Robot Law*. Cheltenham, UK: Edward Elgar Publishing, 3–22.
Rickberg, Merit 2023. Towards Complexity Thinking with Juri Lotman: Modelling Cultural Dynamics in Educational Systems. Dissertationes semioticae Universitatis Tartuensis. Tartu: University of Tartu.
Ritivoi, Andreea Deciu 2009. Explaining People: Narrative and the Study of Identity. *StoryWorlds: A Journal of Narrative Studies* 1: 25–41.
Roberge, Jonathan, and Michael Castelle 2021. Toward an End-to-End Sociology of 21st-Century Machine Learning. In: Roberge, Jonathan; Castelle, Michael (eds.) *The Cultural Life of Machine Learning: An Incursion into Critical AI Studies*. Cham: Springer International Publishing, 1–29.

Ryan, Marie-Laure 1991. *Possible Worlds, Artificial Intelligence, and Narrative Theory*. Bloomington & Indianapolis: Indiana University Press.
Ryle, Gilbert 1949. *The Concept of Mind*. London; New York: Hutchinson's University Library.
Salupere, Silvi 2015. The Cybernetic Layer of Juri Lotman's Metalanguage. *Recherches sémiotiques / Semiotic Inquiry* 35(1): 63–84.
Santosh, K. C., and Casey Wall 2022. *AI, Ethical Issues and Explainability—Applied Biometrics*. Singapore: Springer Nature.
Sap, Maarten, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A. Smith 2019. The Risk of Racial Bias in Hate Speech Detection. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 1668–1678.
Scharre, Paul 2018. *Army of None: Autonomous Weapons and the Future of War*. New York; London: W. W. Norton & Company.
Schemmer, Max, Niklas Kühl, Carina Benz, and Gerhard Satzger 2022. On the Influence of Explainable AI on Automation Bias. arXiv:2204.08859.
Scholem, Gershom 1969. *On the Kabbalah and Its Symbolism*. (Transl. R. Manheim.) New York: Schocken Books.
——— 1978. *Kabbalah*. New York and Scarborough, Ontario: New American Library.
Sebeok, Thomas A. 1979. *The Sign & Its Masters*. Austin: University of Texas Press.
——— 1986. *I Think I Am a Verb: More Contributions to the Doctrine of Signs*. Boston, MA: Springer.
——— 1991. *A Sign Is Just a Sign*. Bloomington: Indiana University Press.
——— 2001. *Signs: An Introduction to Semiotics.* Toronto; Buffalo: University of Toronto Press.
Segal, Howard P. 2012. *Utopias: A Brief History from Ancient Writings to Virtual Communities*. Chichester, West Sussex; Malden, MA: Wiley-Blackwell.
Selg, Peeter, and Andreas Ventsel 2022. Cultural Semiotics as the Foundation of Political Semiotics. *Social Semiotics* 32(5): 671–688.
Shane, Janelle 2019. *You Look like a Thing and I Love You: How Artificial Intelligence Works and Why It's Making the World a Weirder Place*. New York: Voracious/Little, Brown and Company.
Sharkey, Amanda, and Noel Sharkey 2020. We Need to Talk about Deception in Social Robotics! *Ethics and Information Technology* 23: 1–8.
SIA 2016. *Beyond Borders: The Global Semiconductor Value Chain. How an Interconnected Industry Promotes Innovation and Growth*. Semiconductor Industry Association; Nathan Associates.
Silk, Leonard Solomon, and David Vogel 1976. *Ethics and Profits: The Crisis of Confidence in American Business*. New York: Simon and Schuster.
Simpson, Michael, and Alejandra Pizarro Choy 2023. Building Decolonial Climate Justice Movements: Four Tensions. *Dialogues in Human Geography* 20438206231174629.
Singler, Beth 2019. Existential Hope and Existential Despair in Ai Apocalypticism and Transhumanism. *Zygon*® 54(1): 156–176.
——— 2020. "Blessed by the Algorithm": Theistic Conceptions of Artificial Intelligence in Online Discourse. *AI & SOCIETY* 35(4): 945–955.
Sismondo, Sergio 2010. *An Introduction to Science and Technology Studies*. Chichester, West Sussex, U.K.; Malden, MA: Wiley-Blackwell.
Skagestad, Peter 1993. Thinking with Machines: Intelligence Augmentation, Evolutionary Epistemology, and Semiotic. *Journal of Social and Evolutionary Systems* 16(2): 157–180.

——— 1999. Peirce's Inkstand as an External Embodiment of Mind. *Transactions of the Charles S. Peirce Society* 35(3): 551–561.
Smith, Gary 2018. *The AI Delusion*. Oxford: Oxford University Press.
Smith, Linda Tuhiwai 2021. *Decolonizing Methodologies: Research and Indigenous Peoples*. London: Zed.
de Solla Price, Derek J. 1964. Automata and the Origins of Mechanism and Mechanistic Philosophy. *Technology and Culture* 5(1): 9–23.
Solomonoff, Ray 1956. [Notes Describing Certain Days' Activities]. http://raysolomonoff.com/dartmouth/boxa/dartmynotes723815.pdf. Accessed on September 18, 2023.
Strubell, Emma, Ananya Ganesh, and Andrew McCallum 2019. Energy and Policy Considerations for Deep Learning in NLP. arXiv:1906.02243.
Taddeo, Mariarosaria, and Alexander Blanchard 2022. Accepting Moral Responsibility for the Actions of Autonomous Weapons Systems—a Moral Gambit. *Philosophy & Technology* 35(3): 78.
Taylor, Astra 2018. The Automation Charade. *Logic(s) Magazine*. https://logicmag.io/failure/the-automation-charade/. Accessed on September 18, 2023.
Taylor, Trevor 2019. Artificial Intelligence in Defence. *The RUSI Journal* 164(5–6): 72–81.
Thompson, Stith 1961. *The Types of the Folk-Tale, a Classification and Bibliography: Antti Aarne's Verzeichnis der Märchentypen (FF Communications No. 3)*. Helsinki: Suomalainen tiedeakatemia.
Todd, Zoe 2016. An Indigenous Feminist's Take on The Ontological Turn: "Ontology" Is Just Another Word for Colonialism. *Journal of Historical Sociology* 29(1): 4–22.
Tomberg, Jaak 2019. Küberpunk ilma teadusulmeta / Cyberpunk without Science Fiction. *Methis. Studia humaniora Estonica* 18(23): 110–128.
——— 2023. *Kuidas täita soovi. Realism, teadusulme ja utoopiline kujutlusvõime*. Tartu: Tartu Ülikooli Kirjastus.
Tõugu, Enn 2018. *Arvutid, küberruum ja tehismõistus: noppeid arvutite imepärasest eduloost*. Tallinn: Tallinna Ülikooli Kirjastus.
Truitt, Elly Rachel 2004. "Trei Poete, Sages Dotors, Qui Mout Sorent Di Nigromance": Knowledge and Automata in Twelfth-Century French Literature. *Configurations* 12(2): 167–193.
——— 2015. *Medieval Robots*. Philadelphia: University of Pennsylvania Press.
Turing, A. M. 1950. Computing Machinery and Intelligence. *Mind* 59(236): 433–460.
Tybjerg, Karin 2003. Wonder-Making and Philosophical Wonder in Hero of Alexandria. *Studies in History and Philosophy of Science Part A* 34(3): 443–466.
Uther, Hans-Jörg 2011. *The Types of International Folktales. 1: Animal Tales, Tales of Magic, Religious Tales, and Realistic Tales, with an Introduction*. Helsinki: Suomalainen Tiedeakatemia.
Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin 2017. Attention Is All You Need. *arXiv: 1706.03762 [cs]*.
Viidalepp, Auli 2020. Representations of Robots in Science Fiction Film Narratives as Signifiers of Human Identity. *Információs Társadalom* 20(4): 19.
de Visser, Ewart J., Samuel S. Monfort, Ryan McKendrick, Melissa A. B. Smith, Patrick E. McKnight, Frank Krueger, and Raja Parasuraman 2016. Almost Human: Anthropomorphism Increases Trust Resilience in Cognitive Agents. *Journal of Experimental Psychology: Applied* 22(3): 331–349.

de Waal, Frans B. M. 1999. Anthropomorphism and Anthropodenial: Consistency in Our Thinking about Humans and Other Animals. *Philosophical Topics* 27(1): 255–280.
Wallerstein, Immanuel 1974. *Capitalist Agriculture and the Origins of the European World-Economy in the Sixteenth Century*. New York: Academic Press.
——— 2004. *World-Systems Analysis: An Introduction*. Durham and London: Duke University Press.
——— 2006. *European Universalism: The Rhetoric of Power*. New York; London: The New Press.
Wang, Jiawei, and Tianyi Xu 2021. Environmental Impact of Information Communication Technology: A Review of Econometric Assessment Methods, Influential Mechanism, and Influential Direction. *Environmental Impact Assessment Review* 89: 106590.
Watson, David 2019. The Rhetoric and Reality of Anthropomorphism in Artificial Intelligence. *Minds & Machines* 29(3): 417–440.
Wiener, Earl L. 1981. Complacency: Is the Term Useful for Air Safety. In: *Proceedings of the 26th Corporate Aviation Safety Seminar*, 116–125.
Wiener, Earl L., and Renwick E. Curry 1980. Flight-Deck Automation: Promises and Problems. *Ergonomics* 23(10): 995–1011.
Winner, Langdon 1980. Do Artifacts Have Politics? *Daedalus* 109(1): 121–136.
——— 1997. Technology Today: Utopia or Dystopia? *Social Research* 64(3): 989–1017.
Woolgar, Steve 1985. Why Not a Sociology of Machines? The Case of Sociology and Artificial Intelligence. *Sociology* 19(4): 557–572.
van Wynsberghe, Aimee 2021. Sustainable AI: AI for Sustainability and the Sustainability of AI. *AI and Ethics* 1(3): 213–218.
Zeng, Jing, Chung-hong Chan, and Mike S. Schäfer 2020. Contested Chinese Dreams of AI? Public Discourse about Artificial Intelligence on WeChat and People’s Daily Online. *Information, Communication & Society* 25(3): 319–340.
Zhai, Yujia, Jiaqi Yan, Hezhao Zhang, and Wei Lu 2020. Tracing the Evolution of AI: Conceptualization of Artificial Intelligence in Mass Media Discourse. *Information Discovery and Delivery* 48(3): 137–149.
Zhao, Jieyu, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai-Wei Chang 2017. Men Also Like Shopping: Reducing Gender Bias Amplification Using Corpus-Level Constraints. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Copenhagen, Denmark: Association for Computational Linguistics, 2979–2989.
Zhao, Xuan, and Bertram F. Malle 2022. Spontaneous Perspective Taking toward Robots: The Unique Impact of Humanlike Appearance. *Cognition* 224: 105076.
Ziemke, Tom, and Noel E. Sharkey 2001. A Stroll through the Worlds of Robots and Animals: Applying Jakob von Uexkull’s Theory of Meaning to Adaptive Robots and Artificial Life. *Semiotica* 134(1/4): 701–746.
Zuboff, Shoshana 1988. *In the Age of the Smart Machine: The Future of Work and Power*. New York: Basic Books.
——— 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.

SUMMARY IN ESTONIAN

“Oodatav tehisintellekt” kui ühiskondlik-kultuuriline konstrukt ja selle mõju tehnoloogiadiskursusele

Tehisintellekt (TI) on hetkel paljudes aruteludes üks populaarsemaid märksõnu, eriti seoses 2023. aastal alanud võidujooksuga tekstiloomemudelite turule toomises. Praegust tehisintellekti puudutavat avalikku arutelu iseloomustavad eelkõige asjastatud metafoorid, mis kitsendavad tehnoloogia võimalikke tõlgendusi ja tähendusi. See kehtib nii akadeemilise kui päris avaliku diskursuse puhul.

Doktoritöö pakub kriitilise sissevaate tehnoloogiadiskursusse, keskendudes tehisintellekti mõistega seonduvale. Avan töös järgnevaid hüpoteese-mõttekäike:

1. Avalikkuse kujutelmi tehnoloogiast mõjutab palju teadusulme ja -fantastika, samuti intertekstid ja konnotatsioonid folkloorist, religioonist, müütidest. Kõik need kaasnevad diskursused mõjutavad avalikku arusaama tehisintellektist. Samuti kujundavad need ootusi praegustele ja tulevastele tehnoloogilistele lahendustele.
2. Tehisintellekti kui mõiste ja idee kultuuriajaloolised juured moodustavad osa „oodatava tehisintellekti“ kujundist. See on diskursuse eri osadest kujunenud koondtähistaja, millega seonduvad ajaloolised ja ühiskondlik-kultuurilised konnotatsioonid, erinevad tähistusobjektid jne.
3. Tehisintellekti-diskursusel on tugevad antropomorfismi mõjud. Antropomorfism moodustab kaasosa tehnodeterministlikust vaatest, mille järgi tehnoloogia areneb inimese kaasmõjuta. Paljud tehnoloogia loojad toetavad sellist vaadet kas avalikult või alateadlikult ning nad on selleks poliitiliselt motiveeritud.

1. peatükk paigutab uurimuse teoreetilise tausta teadus-ja tehnoloogiauuringute konteksti ning kinnitab TI representatsioonid uurimistöö semiootilise objektina. Mõttekäikude avamiseks ja toestamiseks selgitan kõigepealt peatükis 1.1 semiootika ja tehisintellekti-uuringute ajaloolist ühisosa ning peatükis 1.2 nende mõistelisi seoseid. Näiteks on metakeelse ühisosa loomist taotlenud 1990te masinsemioosi mõiste. Samuti on Uexkülli *Umwelt*-mudelit püütud kasutada arvutiteaduses heuristikana. Semiootikud jäävad enamuses siiski seisukohale, et masinas endas semioosi ei toimu.

2. peatükis keskendun „oodatava tehisintellekti“ mudelile ja selle tähenduste kogumile. Peatükk 2.1 käsitleb TI mõiste ajalugu ja kaasaegseid definitsioone. Peatükis 2.1.1 vaatlen autonoomseid seadeldisi läbi vanakreeka imeasjade poliitika raamistuse. Peatükk 2.2 käsitleb tehnoloogiaga seonduvaid metakeele probleeme, näiteks antropomorfism (2.2.1) ja sellega kaasnevad eelarvamused tehnoloogia toimevõime kohta (2.2.2). Peatükk 2.3 arutleb tehnoloogia väidetava neutraalsuse üle ning küsib, kellel on tänases ühiskonnas üldse lubatud tehnoloogiast kõnelda.

Doktoritöö pakub välja ka eetilisema ja terviklikuma ontoloogilise mudeli TI-põhiste süsteemide kirjeldamiseks. Mudel kirjeldab süsteeme kui kompleksseid koosluseid, arvestades nende ühiskondlik-materiaalset korrastatust, üleilmset majanduslikku ja materiaalset kujunemist. Samuti arvestab mudel TI-süsteemide mõju keskkonnale, ühiskondlikele institutsioonidele ja semiosfäärile. Doktoritöö näitab, et tehisintellekti ei tuleks mõista mitte ainult kui objekti või sotsiotehnilist süsteemi oma kultuuriajalooliste juurtega, vaid kui kogu tooteahela tervikut, mis hõlmab inimesi, kultuure, kasutatud ressursse ning nii materiaalset kui semiootilist mõju terve planeedi tasandil.

Doktoritöö osaks on kolm artiklit, mis käsitlevad erinevaid TI representatsioone, generatiivse TI-ga loodud sünteetilist sisu ja selle ühiskondlikke toimimispõhimõtteid.

I artikkel analüüsib viimase kümnendi ulmefilmiseeriate robottegelasi. Artikkel käsitleb roboteid osana üliinimeste spektrist, kuhu kuuluvad muidu igasugused ülivõimetega tegelaskujud, nt küborgid, superkangelased ja ka tehisolendid. Artiklis analüüsitakse tehisolendeid läbi semioosi ja ennustamatuse mõistete ning leitakse, et robottegelane tähistab tekstivälise reaalsuse osas ulmežanrile tavapärasena siiski inimest ennast läbi tema võõrandamise ühiskonnast. Lisaks peegeldab robottegelane ühiskondlikku arusaama (hullust) teadlasest ja tema tööeetikast.

II artikkel käsitleb generatiivtekstide sotsiokommunikatsiivseid funktsioone (Lotmani järgi). Analüüsitakse 2020.a septembris ajalehes *the Guardian* avaldatud GPT-3 tekstiloomemudeli abiga genereeritud teksti, mis avaldati „arvamusartikli“ nime all, ning selle retseptsiooni uue meedia väljaannetes. Nimelt mitmed väljaanded kommenteerisid ja tsiteerisid *the Guardiani* artiklit kui „roboti arvamust“, mis tekitas üksjagu poleemikat. Artikkel püüab Lotmani tekstifunktsioonide mudeli abil avada antud generatiivteksti võimalikke tõlgendusi ning põhjusi, miks seda võidakse tajuda sõnumina „robot-autorilt“. Artikkel näitab ka, kuidas sünteetilise teksti puhul autori mõiste hägustub ning autorlus paigutub laiali erinevate osapoolte vahel, kelleks on toimetajad, mudeliga töötajad ning ka teksti lõpplugeja.

III artikkel arutleb generatiivmeedia üle üldisemal ja teoreetilisemal tasandil. Süvavõltsing-videote ja generatiivtekstide näitel küsitakse, kus asub antud kultuuritekstide viitereaalsus ja kas seda üldse on. Mimikri ja nonsensi mõistete abil näitab artikkel, kuidas generatiivmeedia puhul tavapärase kommunikatsiooniakti osapooled muutuvad või on esmapilgul lugeja (vastuvõtja) eest varjatud.

PUBLICATIONS

CURRICULUM VITAE

Name:	Auli Viidalepp
Address:	Department of Semiotics, Jakobi 2–318, 51005 Tartu
E-mail:	auli.viidalepp@ut.ee
Homepage:	technosemiotics.net

Education

2019–2023	University of Tartu, PhD, Semiotics
2016–2019	University of Tartu, MA, Semiotics, *cum laude*
2000–2004	University of Tartu, BA, Semiotics

Study visits at other universities

2022/05–2023/07	Univerzita Palackého v Olomouci
2018/02–07	Eötvös Loránd Tudományegyetem
2017/10–12	Université Bordeaux Montaigne

Academic career

2021/09–2022/05	Estonian Military Academy, researcher, EDIDP project iMUGS
2021/01–08	University of Tartu, Department of Semiotics, Junior Researcher

Research-related managerial and administrative work

2020–2024	EU COST Action CA19102 LITHME, Grant Awarding Coordinator
2023/08–…	ICOHTEC, member
2023/08–…	International Semiotics Institute, member
2023/01–…	Technosemiotics research network, founder
2004–…	Estonian Semiotics Association, member

Research topics:
technosemiotics, discourse on technology, science and technology studies

ELULOOKIRJELDUS

Nimi:	Auli Viidalepp
Aadress:	TÜ semiootika osakond, Jakobi 2–318, 51005 Tartu
E-post:	auli.viidalepp@ut.ee
Koduleht:	technosemiotics.net

Haridus

2019–2023	Tartu Ülikool, doktoriõpe, Semiootika ja kultuuriteooria
2016–2019	Tartu Ülikool, magistriõpe, Semiootika ja kultuuriteooria, *cum laude*
2000–2004	Tartu Ülikool, bakalaureuseõpe, Semiootika ja kultuuriteooria

Õppetöö teistes ülikoolides

2022/05–2023/07	Univerzita Palackého v Olomouci
2018/02–07	Eötvös Loránd Tudományegyetem
2017/10–12	Université Bordeaux Montaigne

Teenistuskäik

2021/09–2022/05	Kaitseväe Akadeemia, teadustöötaja EDIDP projektis iMUGS
2021/01–08	Tartu Ülikool, Semiootika osakond, Nooremteadur

Muu teadusorganisatsiooniline ja -administratiivne tegevus

2020–2024	EU COST Action CA19102 LITHME, liige
2023/08–…	ICOHTEC, liige
2023/08–…	International Semiotics Institute, liige
2023/01–…	Tehnosemiootika uurimisvõrgustik, asutaja
2004–…	Eesti Semiootika Selts, liige

Uurimisvaldkonnad:
tehnosemiootika, tehnoloogiadiskursus, teadus- ja tehnoloogiauuringud

DISSERTATIONES SEMIOTICAE UNIVERSITATIS TARTUENSIS

1. **Михаил Ю. Лотман.** Структура и типология русского стиха. Тарту, 2000.
2. **Елена Григорьева.** Эмблема: структура и прагматика. Тарту, 2000.
3. **Valdur Mikita.** Kreatiivsuskäsitluste võrdlus semiootikas ja psühholoogias. Tartu, 2000.
4. **Ирина Аврамец.** Поэтика новеллы Достоевского. Тарту, 2001.
5. **Ян Левченко.** История и фикция в текстах В. Шкловского и Б. Эйхенбаума в 1920-е гг. Тарту, 2003.
6. **Anti Randviir.** Mapping the world: towards a sociosemiotic approach to culture. Tartu, 2004.
7. **Timo Maran.** Mimikri kui kommunikatsiooni-semiootiline fenomen. Tartu, 2005.
8. **Элеонора Рудаковская-Борисова.** Семиотика пищи в произведениях Андрея Платонова. Tartu, 2005.
9. **Andres Luure.** Duality and sextets: a new structure of categories. Tartu, 2006.
10. **Peeter Linnap.** Fotoloogia. Tartu, 2006.
11. **Daniele Monticelli.** Wholeness and its remainders: theoretical procedures of totalization and detotalization in semiotics, philosophy and politics. Tartu, 2008.
12. **Andreas Ventsel.** Towards semiotic theory of hegemony. Tartu, 2009.
13. **Elin Sütiste.** Tõlke mõiste dünaamikast tõlketeaduses ja eesti tõlkeloos. Tartu, 2009.
14. **Renata Sõukand.** Herbal landscape. Tartu, 2010.
15. **Kati Lindström.** Delineating Landscape Semiotics: Towards the Semiotic Study of Landscape Processes. Tartu, 2011.
16. **Morten Tønnessen.** Umwelt transition and Uexküllian phenomenology. An ecosemiotic analysis of Norwegian wolf management. Tartu, 2011. 231 p.
17. **Anu Sarv.** Õppejõu eneserefleksioon ja professionaalne identiteet. Tartu, 2013, 141 lk.
18. **Vadim Verenich.** Semiotic models of legel argumentation. Tartu, 2014, 249 lk.
19. **Maarja Ojamaa.** The transmedial aspect of cultural autocommunication. Tartu, 2015, 173 p.
20. **Tiit Remm.** Sociocultural Space: Spatial Modelling and the Sociocultural World. Tartu, 2015, 142 p.
21. **Riin Magnus.** The Semiotic Grounds of Animal Assistance: Sign Use of Guide Dogs and Their Visually Impaired Handlers. Tartu, 2015, 154 p.
22. **Davide Weible.** Exaptation: Towards a Semiotic Account of a Biological Phenomenon. Tartu, 2016, 144 p.

23. **Mari-Liis Madisson.** The Semiotic Construction of Identities in Hypermedia Environments: The Analysis of Online Communication of the Estonian Extreme Right. Tartu, 2016, 195 p.
24. **Claudio Julio Rodríguez Higuera.** The Place of Semantics in Biosemiotics: Conceptualization of a Minimal Model of Semiosic Capabilities. Tartu, 2016, 148 p.
25. **Kadri Tüür.** Semiotics of Nature Representations: on the Example of Nature Writing. Tartu, 2017, 257 p.
26. **Gleb Netchvolodov.** Pictorially simplified images as machine vision design tool: semiotics approach. Tartu, 2017, 143 p.
27. **Сильви Салупере.** О метаязыке Юрия Лотмана: проблемы, контекст, источники. Tartu, 2017, 210 c.
28. **Remo Gramigna.** Augustine and the study of signs and signification. Tartu, 2018, 237 p.
29. **Nelly Mäekivi.** The Zoological Garden as a Hybrid Environment – A (Zoo)semiotic Analysis. Tartu, 2018, 172 p.
30. **Silver Rattasepp.** The Human Mirror. A Critique of the Philosophical Discourse on Animals from the Position of Multispecies Semiotics. Tartu, 2018, 156 p.
31. **Ott Puumeister.** On Biopolitical Subjectivity: Michel Foucault's perspective on biopolitics and its semiotic aspects. Tartu, 2018, 170 p.
32. **Lyudmyla Zaporozhtseva.** Structural Units of Mass Culture Mythology: A Cultural Semiotic Approach. Tartu, 2019, 193 p.
33. **Tatjana Menise.** Fairy Tales in Transmedia Communication: Fanfiction. Tartu, 2020, 142 p.
34. **Alexandra Milyakina.** Digitalization of Literary Education in the Context of Cultural Autocommunication. Tartu, 2020, 137 p.
35. **Tyler James Bennett.** Detotalization and retroactivity: black pyramid semiotics. Tartu, 2021, 238 p.
36. **Jason Mario Dydynski.** Semiotic Modeling of Cuteness in Cartoon Characters/Mascots. Tartu, 2021, 176 p.
37. **Tiina Hoffmann.** La traduction cinématographique en Estonie soviétique : contextes, pratiques et acteurs. Tartu, 2021, 489 p.
38. **Mirko Cerrone.** Interspecies Relationships: A zoosemiotic Analysis of Human-Ape Communication. Tartu, 2021, 160 p.
39. **Jonathan Grant Griffin.** A Semiotic of Motives: Kenneth Burke and Deely-Tartu Semiotics. Tartu, 2022, 244 p.
40. **Kyle Jonathan Davidson**. Semiotic Modelling of Identity and Communication in Virtual Reality, Augmented Reality, and Mixed Reality. Tartu, 2022, 221 p.
41. **Aleksandr Fadeev.** Lev Vygotsky's approach in the Contemporary Semiotic Research of Learning, Meaning-Making and Inner Speech. Tartu, 2022, 146 p.
42. **Eugenio Israel Chávez Barreto.** The Semiotic Theory of Luis Jorge Prieto. Tartu, 2022, 236 p.

43. **Katarina Damčević.** Semiotics of Hate Speech and Contested Symbols: The “Za dom spremni” Ustaša Salute in Contemporary Croatia. Tartu, 2023, 166 p.
44. **Merit Rickberg.** Towards Complexity Thinking With Juri Lotman: Modelling Cultural Dynamics in Educational Systems. Tartu, 2023, 233 p.
45. **Tuuli Pern**. Affective-imaginative modelling in semiotic context: A Vichian perspective. Tartu, 2023, 147 p.