THE HUMAN CONDITION IN AN ALGORITHMIZED WORLD

A CRITIQUE THROUGH THE LENS OF 20TH-CENTURY JEWISH THINKERS AND THE CONCEPTS OF RATIONALITY, ALTERITY AND HISTORY

Nathalie A. Smuha*

This paper can be cited as:
Nathalie A. Smuha (2022), "The human condition in an algorithmized world: a critique through the lens of 20th-century Jewish thinkers and the concepts of rationality, alterity and history", Institute of Philosophy, KU Leuven.

ABSTRACT

Artificial Intelligence (AI) systems are increasingly deployed in all domains of our lives. While their use can provide substantial benefits, they also entail significant risks – and ethics has been put forward as the key solution to counter these risks. Yet the manner in which ethics is typically relied on in this context is woefully deficient. At best, ethics is given the role of orienting problematic technology towards 'acceptable' uses, thereby legitimizing AI's widespread adoption, which is taken for granted. At worst, ethics is instrumentalized as a quality-label to stimulate AI's deployment, as part of a broader doctrine of 'progress'. Current ethics discourse hence appears unable to provide a more fundamental critique of the way in which the algorithmized world is profoundly impacting our existence. This is because it starts from within a technological paradigm that does not fundamentally question AI's place and progression in society. In this paper, I therefore argue that, if ethics is to shed light on – and to offer a more fundamental critique of – the human condition in an algorithmized world, without being bound to today's technological paradigm, it requires a meta-technological perspective that puts ethics first. To pursue this aim, I propose to ground our approach in the fact that our existence in the world is necessarily intersubjective and relational, and use the lens of intersubjectivity to examine AI's impact on the human condition. To narrow the scope of my analysis, I focus on AI's impact on three interrelated domains of our existence: (1) our way of thinking or rationality, (2) our way of engaging with others or alterity and (3) our way of experiencing time or history. In my analysis, I draw on the work of 20th-century Jewish thinkers, such as Franz Rosenzweig, Emmanuel Levinas and Hannah Arendt, given the importance they ascribe to relationality and its role in countering totalitarian thinking which, as I argue, can also arise through the systemic irresponsible use of AI.

After introducing my research inquiry (Chapter 1) and providing a brief definition of AI (Chapter 2), I seek to answer three questions: First, what does the algorithmized world look like, and what is its underpinning societal paradigm (Chapter 3)? Second, how does current AI ethics discourse approach AI's risks, and how does it fall short of delivering a more fundamental critique of AI's impact on the human condition (Chapter 4)? Third, how does AI's ubiquity affect the human condition, and particularly our experience of rationality, alterity and history (Chapter 5)? Based on my research findings, I conclude that the totalizing use of AI systems – and the way it impacts our way of thinking, our way of engaging with others and our way of experiencing time – can give rise to significant concerns, as it may be used in a way that opposes our ability to live a meaningful life by engaging in intersubjective human relationships. To close this paper, I postulate several avenues that should be explored to counter the concerns identified (Chapter 6).

---

* Researcher, KU Leuven Faculty of Law, Tiensestraat 41, 3000 Leuven, nathalie.smuha@kuleuven.be.

I wish to express my gratitude to Roger Vergauwen and Luc Anckaert for their invaluable support in the context of this research, which I conducted during my Philosophy degree (MA) at the KU Leuven Institute of Philosophy.

TABLE OF CONTENTS

1. INTRODUCTION

Artificial Intelligence (AI) systems are increasingly deployed in all domains of our lives. While their use can provide substantial benefits, they also entail significant risks – from perpetuating discriminatory practices to enabling mass-surveillance.[1] To counter these problems and "*maximize the benefits of AI systems while at the same time preventing and minimizing their risks*"[2], ethics has been put forward as the key solution. From practical guidelines to ensure that AI is 'ethical-by-design', to proposals for new legislation to ensure 'ethical AI', over the past years, ethics discourse has permeated the technological realm and gained an ever more prominent role therein.[3]

The manner in which ethics is typically relied on in this context is, however, woefully deficient.[4] At best, ethics is given the role of orienting problematic technology towards 'acceptable' uses, thereby simultaneously legitimizing AI's widespread adoption, which is taken for granted. At worst, ethics is instrumentalized as a quality-label to stimulate AI's deployment, as part of a broader doctrine of 'progress', grounded in meliorism. Either way, the current approach to ethics in the sphere of AI appears unable to deliver a more fundamental critique of AI's adverse impact. This is because it starts from within a paradigm that does not fundamentally question AI's place and progression in society. Due to this deficit, much of the contemporary AI ethics discourse is only scratching the surface of the ways in which this technology can alter human existence. Indeed, beyond problems of biased data and faulty design, the scaled and cumulative use of AI risks profoundly affecting our being-in-the-world.

---

1 Corinne Cath et al., 'Artificial Intelligence and the "Good Society": The US, EU, and UK Approach', *Science and Engineering Ethics*, 28 March 2017; Cathy O'Neil, *Weapons of Math Destruction* (Penguin Books Ltd, 2017); Emre Bayamlıoğlu and Ronald Leenes, 'The "Rule of Law" Implications of Data-Driven Decision-Making: A Techno-Regulatory Perspective', *Law, Innovation and Technology* 10, no. 2 (3 July 2018): 295–313; Joy Buolamwini and Timnit Gebru, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification', in *Proceedings of Machine Learning Research*, vol. 81, 2018, 1–15, http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf; Karen Yeung, 'Why Worry about Decision-Making by Machine?', in *Algorithmic Regulation*, ed. Karen Yeung and Martin Lodge (Oxford University Press, 2019), 21–48; Kate Crawford et al., *AI Now 2019 Report* (New York: AI Now Institute, 2019), https://ainowinstitute.org/AI_Now_2019_Report.pdf; M. Brkan, 'Artificial Intelligence and Democracy', *Delphi - Interdisciplinary Review of Emerging Technologies* 2, no. 2 (2019): 66–71.

2 High-Level Expert Group on AI, 'Ethics Guidelines for Trustworthy AI', 8 April 2019.

3 Besides the above-cited Ethics Guidelines for Trustworthy AI, see also Mike Ananny, 'Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness', *Science, Technology, & Human Values* 41, no. 1 (January 2016): 93–117; Paula Boddington, *Towards a Code of Ethics for Artificial Intelligence*, Artificial Intelligence: Foundations, Theory, and Algorithms (Cham: Springer International Publishing, 2017); Anna Jobin, Marcello Ienca, and Effy Vayena, 'The Global Landscape of AI Ethics Guidelines', *Nat Mach Intell* 1 (2019): 389–99; Thilo Hagendorff, 'The Ethics of AI Ethics: An Evaluation of Guidelines', *Minds and Machines* 30, no. 1 (March 2020): 99–120.

4 As will be explained under Chapter 4, criticism on ethics discourse in the context of AI is not new, yet is often limited to the need for *binding* legal rules as opposed to a mere *voluntary* approach. See, for instance, Ben Wagner, 'Ethics As An Escape From Regulation. From "Ethics-Washing" To Ethics-Shopping?', in *Being Profiled*, ed. Emre Bayamlioglu et al. (Amsterdam: Amsterdam University Press, 2019), 84–89; Elettra Bietti, 'From Ethics Washing to Ethics Bashing', *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, January 2020, 210–19; Karen Yeung, Andrew Howes, and Ganna Pogrebna, 'AI Governance by Human Rights–Centered Design, Deliberation, and Oversight: An End to Ethics Washing', in *The Oxford Handbook of Ethics of AI*, ed. Markus D. Dubber, Frank Pasquale, and Sunit Das (Oxford University Press, 2020), 75–106.

If ethics is to shed light on – and to offer a more fundamental critique of – the human condition in an algorithmized world, without being bound to today's technological paradigm, it requires a third, meta-technological approach[5] that starts from an Archimedean point.[6] Without claiming that we can or should entirely abstract ourselves from today's technological reality, it is by transcending its historicity and finding a reference point outside the technology that we can have a more holistic perspective of the current technological developments and their consequences for our being-in-the-world. Rather than starting from the technology and moving to ethics, I therefore propose to start our inquiry the other way around, hoping that this increases our chances to avoid the pitfalls of the current discourse.

For the purpose of this paper, I seek this Archimedean point in the undeniable meta-technological fact that our 'being-in-the-world' is essentially a 'being-in-the-world-together', and that this intersubjective relationality is an essential characteristic of human existence. To develop this point, and to assess how the widespread use of AI is changing the human condition, I draw on the work of twentieth-century Jewish thinkers[7], and most prominently the writings of Franz Rosenzweig, Emmanuel Levinas and Hannah Arendt. The reason for this is threefold.

First, Judaism assigns a central role to relationality[8], and the importance thereof has been explored in great detail by Jewish thinkers.[9] Second, the horrors of the First and Second World War resulted in remarkable contributions by Jewish authors that conceptualize the human condition as essentially intersubjective. Third, each of the aforementioned authors also contributed to the philosophical understanding of totalitarian thinking, and how this affects our intersubjective relationships. These insights are particularly relevant in the context of AI, given the analogies that can be drawn between certain problematic uses of AI on the one hand, and the exacerbation of totalitarian practices on the other hand.[10]

Indeed, much like totalitarian tactics, AI systems can – consciously or unconsciously, by error or by design – be used to polarize and dehumanize people, undermine their sense of judgment and accountability, and hollow out the responsibilities that accompany our being-in-the-world-together. These effects run counter to respecting the intersubjective nature of our existence.

---

5 Inspiration is drawn in particular from Luc Anckaert's examination of the role of ethics in the context of Globalisation, in Luc Anckaert, 'Globalisation and the Tragedy of Ethics', in *Building Towers: Perspectives on Globalisation*, ed. Luc Anckaert, Danny Cassimon, and Hendrik Opdebeeck, Ethical Perspectives Monograph Series 2 (Leuven: Peeters, 2002), 9–36.

6 An Archimedean point or *Punctum Archimedis* refers to the alleged statement by Archimedes that he could lift the Earth off its foundation if only he were given a solid place to stand and a long enough lever. It has been referred to by Descartes in his *Meditations*, when he sought to find a point that is certain and indubitable.

7 Given the space limitations of this paper, rather than providing a comprehensive overview of the writings of the Jewish authors I discuss, I solely focus on selected insights that are of relevance to this paper's subject.

8 Judaism is, of course, not the only religion of which this can be said.

9 I use the term 'thinkers' or 'authors' rather than 'philosophers', given that Hannah Arendt, for instance, did not consider herself as a philosopher but rather as a political theorist. See e.g. Steve Buckler, *Hannah Arendt and Political Theory: Challenging the Tradition* (Edinburgh University Press, 2011).

10 Larry Diamond, 'The Threat of Postmodern Totalitarianism', *Journal of Democracy* 30, no. 1 (2019): 20–24; Di Minardi, 'The Grim Fate That Could Be "Worse than Extinction"', *BBC*, 16 October 2020, https://www.bbc.com/future/article/20201014-totalitarian-world-in-chains-artificial-intelligence; Nathalie A. Smuha, 'Trustworthy Artificial Intelligence in Education: Pitfalls and Pathways' (Social Science Research Network, 2020).

Accordingly, this paper hypothesizes that the insights of Jewish thinkers that are relevant when examining the risks of totalitarianism, can also be relevant when analyzing the impact of AI.

In light of the above, my aim for this paper is to provide a fundamental critique of the human condition in an algorithmized world by taking an Archimedean perspective of ethics grounded in the inherent intersubjectivity of human existence. To narrow down the scope of my analysis, I focus on AI's impact on three central domains of our being – (1) our way of thinking or *rationality*, (2) our way of engaging with others or *alterity* and (3) our way of experiencing time or *history*.[11] These three domains are closely entwined and, while they can be analyzed in a distinct manner, their relationship to each other corresponds to a ternary structure – which is constitutive of much of Jewish existentialism.[12]

Drawing on the writings of Jewish thinkers, I seek to examine the way in which AI systems infuse these domains with an inherently different logic than the intersubjective one, and what the consequences thereof are for our being-in-the-world-together.[13] In conclusion, based on an analogy between AI's adverse effects and the risks of totalitarianism, I build on these authors' insights to formulate potential avenues that can help counter these effects, and suggest that those avenues merit further exploration in future research.

To pursue this aim, I center this paper around three research questions:

(1) How can the algorithmized world be conceptualized, and what is its underpinning paradigm?

In Chapter 3 of this paper, I conceptualize the algorithmized world and describe its characteristics. Furthermore, I pay particular attention to the paradigm that enables AI's ubiquity, and examine the validity of its underlying assumptions. I conclude that this paradigm raises a number of concerns, which lay at the basis of AI's potential to cause adverse effects.

(2) How does current ethics discourse approach these issues and in which ways is it falling short of delivering a more fundamental critique of AI's impact on our being-in-the-world?

In Chapter 4 of this paper, I provide an overview of the way in which ethics is currently deployed to examine and tackle the problems raised by AI, starting with the rise of ethics guidelines and culminating in the translation of such guidelines into binding legislation. I then explain how this approach is deficient, and how the move to a meta-technological discourse might remedy this deficiency, grounded in the Archimedean point of human intersubjectivity.

---

[11] Though unrelated to the context of AI, these domains have also been analyzed through the lens of Jewish philosophy by Luc Anckaert, whose philosophical analysis of Franz Rosenzweig and Emmanuel Levinas in particular provided inspiration for this work. See e.g. Luc Anckaert, *A Critique of Infinity: Rosenzweig and Levinas*, Studies in Philosophical Theology 35 (Leuven: Peeters, 2006).

[12] Jewish thinkers like Frans Rosenzweig (as well as Martin Buber) greatly relied on the primacy of ternary relationships (over binary relationships) in their contributions. See in this regard, for instance, Luc Anckaert, *God, wereld en mens: het ternaire denken van Franz Rosenzweig*, Wijsgerige verkenningen 17 (Leuven: Universitaire Pers Leuven, 1997).

[13] Note how this ternary structure stands in stark opposition to the binary structure that is inherent to information systems, including AI systems. See in this regard also Dany-Robert Dufour, *Les mystères de la trinité*, Bibliothèque des sciences humaines (Paris: Gallimard, 1990). I dwell further on this point in Chapter 5.1 below.

(3) How does the impact of AI systems affect the human condition – and in particular our experience of rationality, alterity and history?

In Chapter 5 of this paper, I consider how the way we think, the way we engage with others and the way we experience time is being altered by the ubiquitous deployment of AI. I examine how the risks of AI can correlate with the risks engendered by totalitarian systems, and thus how the algorithmized world can exacerbate totalizing forces in society.

Finally, in Chapter 6, I draw on my research findings to provide concluding remarks and postulate a number of avenues that should be explored if we wish to defy the concerns identified. Yet before we can start with this paper's inquiry, a note should be made about Artificial Intelligence – the algorithm-based technology that is central thereto. Accordingly, in the next Chapter, I first define AI for the purpose of this paper.

2. ARTIFICIAL INTELLIGENCE

The term 'Artificial Intelligence' was coined by John McCarthy in 1956[14], six years after Alan Turing's seminal paper kickstarted a broader philosophical discussion of AI with the question: 'Can Machines Think?'.[15] While AI has known several summers – during which scientific breakthroughs gave AI research a boost – and several winters – during which its hyped expectations did not match reality – the technology continuously evolved and gained terrain. Today, AI is enjoying a summer to envy.[16] Many of AI's recent successes can be attributed to data-driven AI systems as opposed to the more traditional code-driven systems[17], since the former are particularly benefitting from the availability of big data and big computing power – propelling the adoption of this algorithm-based technology in ever more domains. Thanks to these successes, the topic of AI is surfing on a wave of attention.

At the same time, until today no uniformly accepted definition of AI exists.[18] Definitions of AI – and the range of technological applications that fall under it – tend to depend on the purpose

---

[14] The conference proposal, submitted in 1955, also detailed the various subjects and methods that would be covered. See John McCarthy et al., 'A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence', 31 August 1955, http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf.

[15] Alan Turing, 'Computing Machinery and Intelligence', *Mind* 59, no. 236 (October 1950): 433–60.

[16] Nathalie A. Smuha, 'Beyond a Human Rights-Based Approach to AI Governance: Promise, Pitfalls, Plea', *Philosophy & Technology*, 24 May 2020.

[17] A distinction is often made between traditional or code-driven AI systems on the one hand, and data-driven or learning-based AI systems on the other hand. The former cover techniques that rely primarily on the codification of symbols and rules, based on which the system 'reasons' (using a top-down approach to design the system's behavior). The latter cover techniques that rely primarily on large amounts of data, based on which the system 'learns' by identifying patterns and creating its own model (using a bottom-up approach to design the system's behavior). The distinction between both should however not be seen as strict; models can be hybrid and incorporate elements of both techniques. See also Virginia Dignum, *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*, Artificial Intelligence: Foundations, Theory, and Algorithms (Springer International Publishing, 2019); Mireille Hildebrandt, 'Algorithmic Regulation and the Rule of Law', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, no. 2128 (13 September 2018).

[18] See Miriam C. Buiten, 'Towards Intelligent Regulation of Artificial Intelligence', *European Journal of Risk Regulation* 10, no. 1 (March 2019): 41–59; Council of Europe Ad Hoc Committee on Artificial Intelligence (CAHAI), 'Feasibility Study' (Strasbourg: Council of Europe, 17 December 2020), https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da; Nathalie A. Smuha, 'From a "Race to AI" to a "Race to AI

of the definition in question. Yet the trend towards global governance initiatives for AI – including by supranational and intergovernmental organizations[19] – has been forcing experts to enhance consensus on how AI should be defined, and what its distinctive characteristics are. Of note is in particular the definition provided by the European Commission's High-Level Expert Group on AI, published alongside its *Ethics Guidelines for Trustworthy AI*[20] in April 2019:

> Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans[21] that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions.[22]

The following elements can already be distilled from this definition. First, these systems are designed by human beings. This means that they do not 'overcome' us passively, but are an active creation of humans, who are thus also responsible for the consequences thereof.[23] Second, there is no 'single' AI system.[24] Rather, AI is an umbrella term for a range of algorithmic technologies that have certain properties in common, namely their ability to reason on or learn from the data provided to them, and to "*act in the physical or digital dimension*" on the basis of such data. This includes applications as diverse as voice recognition systems like Apple's Siri, language processing systems like Google Translate, shopping recommender systems used by Amazon, or self-driving vehicles like robo-taxis. Accordingly, unless explicitly stated otherwise, the term 'AI' as used throughout this paper denotes AI applications more generally rather than one specific AI technology. Third, AI systems can adapt their behavior over time[25] based on what they 'learn'.[26]

Regulation": Regulatory Competition for Artificial Intelligence', *Law, Innovation and Technology* 13, no. 1 (2 January 2021): 57–84.

19 The European Union, as well as the OECD, the Council of Europe and UNESCO, are for instance all in the process of promulgating binding and non-binding legal instruments on AI.

20 High-Level Expert Group on AI, 'Ethics Guidelines for Trustworthy AI'. See also Nathalie A. Smuha, 'The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence', *Computer Law Review International* 20, no. 4 (2019): 97–106.

21 The footnote in the definition states that: "*Humans design AI systems directly, but they may also use AI techniques to optimize their design*." It was added by the Expert Group to reflect the fact that AI systems can sometimes also be programmed to develop new algorithms.

22 High-Level Expert Group on AI, 'A Definition of AI: Main Capabilities and Scientific Disciplines', 8 April 2019.

23 Nathalie A. Smuha, 'Laten We Intelligenter Zijn Wanneer We Het over Artificiële Intelligentie Hebben', Knack Data News, 11 March 2020, https://datanews.knack.be/ict/nieuws/laten-we-intelligenter-zijn-wanneer-we-het-over-artificiele-intelligentie-hebben/article-opinion-1574905.html.

24 Smuha, 'From a "Race to AI" to a "Race to AI Regulation"'.

25 They can do this in an autonomous manner, yet only once they are programmed to do so by a human being.

26 Besides formal definitions provided by governmental organizations, consider also the definition(s) provided by Russell and Norvig in their influential Handbook on AI: Stuart Jonathan Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, Fourth edition, Pearson Series in Artificial Intelligence (Hoboken: Pearson, 2021).

While the European Commission endorsed the High-Level Expert Group's *Ethics Guidelines*, it did not retain the above definition in its subsequent proposal for a new AI regulation, published two years later in April 2021. The reason for this is, presumably, the definition's length and openness, which could be considered as not conducive to legal certainty. Hence, the Commission suggested a somewhat different definition of AI:

> 'Artificial intelligence system' means software that is developed with one or more of the techniques and approaches listed in Annex I[27] and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.[28]

This definition has the advantage of being closer to the one proposed by the OECD[29] and is thus seemingly more reflective of the slowly growing global definitional consensus. Compared with the first definition, it provides us with both more and less information. It provides us with less, since it focuses primarily on 'software' and omits references to different AI techniques – which are instead demoted to a list in Annex I of the proposed regulation. It provides us with more, by concretizing the 'actions' that AI systems can be programmed to undertake based on the data they are fed with, like making recommendations and predictions. Interestingly, the fact that the objectives of AI are 'human-defined' is a prominent element of both definitions.[30]

With these points in the back of our minds, we can conclude this definitional chapter[31] by summing up the main properties associated with AI systems: their ability to reason and learn autonomously based on a purpose that was defined by human beings, their ability to act on that basis in the physical or digital world, and their ability to adapt to their environment and evolve over time, based on new learnings. We can add two more abilities which are omitted from these definitions given their trivial nature, yet which nevertheless merit being rendered explicit here:

---

27 Annex I of the proposed regulation lists the following 'techniques and approaches': (a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning; (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems; (c) Statistical approaches, Bayesian estimation, search and optimization methods.

28 European Commission, 'Proposal for a Regulation of the European Parliament and the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts.', Pub. L. No. COM(2021) 206 final, 2021/0106 (COD) (2021). It should be noted that this proposal is currently being negotiated by the European Parliament and Council, and that this proposed definition can still change during the negotiation rounds prior to the regulation's adoption.

29 In its policy documents, the OECD defines AI systems as follows: "An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy." See OECD, 'Recommendation of the Council on Artificial Intelligence'.

30 This point can be criticized, since AI systems can also be programmed to set objectives autonomously (in light of certain restraints and/or elements of information provided to them). Some hence argue that these systems might fall outside of the scope of the Commission's AI definition (and hence of the future regulation). However, one can counter-argue that, even for those systems, there is still a programming phase during which a human being sets out the system's objectives on a more abstract level, to be further concretized by the system later on, and hence that they do fall under the definition's scope.

31 These definitions (like most AI definitions) focus primarily on AI's technical aspects. Some authors thus started to provide broader definitions of AI, to emphasize its embeddedness in social practices, politics and culture. Consider, for instance, Kate Crawford, *Atlas of AI* (New Haven: Yale University Press, 2021), 8. I elaborate more on the societal aspects of AI in Chapter 3.

their ability to process vast amounts of data at a significant speed, and their ability to operate on a very large scale.[32] Many of the risks arising from the use of AI can also manifest themselves through the use of other technologies, including more basic IT systems that are not necessarily considered 'intelligent' or sophisticated enough to be called AI. Yet the distinctive features of AI systems – and particularly their scale, speed and 'autonomy' – are able to exacerbate those risks, which resulted in technology-specific attention to AI's concerns.

Finally, it should be stressed that the risks arising from AI not only depend on the AI system itself, but also on the particular context in which it is being developed and used – as will be discussed more thoroughly in the next Chapter.[33] More specifically, my focus in this paper concerns AI systems that can have an adverse impact on the interests and rights of individuals, collectives and societies, directly or indirectly, for instance by causing harms or wrongs. Hence, while I certainly acknowledge that AI systems could be developed and deployed in ways and contexts that need not be problematic, this paper particularly focuses on where it can go wrong. With this caveat in mind, we can now further pursue the aim of this paper, starting with an outline of the algorithmized world.

3. THE ALGORITHMIZED WORLD

Preliminary to our investigation of what it means to be human in an 'algorithmized world', let me first elucidate what I mean with this concept. What does an algorithmized world look like, and what renders it so distinctive that it merits an investigation? I intend to answer this question by setting out the emblematic characteristics of such a world, which I group into three sections: the good (3.1), the bad (3.2) and the ugly (3.3).

3.1 The good

In the second sentence of this paper's introduction, I stated that AI systems can "*provide substantive benefits*", after which the rest of the text thus far exclusively focused on AI's risks. While a thorough examination of the benefits of AI falls outside the scope of this paper, it is nevertheless worth highlighting their existence to clarify why organizations are relying on this technology. After all, if AI systems only carried adverse effects, the incentives to embed the world with this technology would be far less present.

In this regard, it is important to stress that the progress of the algorithmized world which we are witnessing today primarily stems from the desire to reap the benefits of the technology's use, with the – often explicit – aim to improve individual and societal welfare.[34] Precisely this desire,

---

[32] Pekka Ala-Pietilä and Nathalie A. Smuha, 'A Framework for Global Cooperation on Artificial Intelligence and Its Governance', in *Reflections on Artificial Intelligence for Humanity*, ed. Bertrand Braunschweig and Malik Ghallab (Cham: Springer International Publishing, 2021), 237–65.

[33] The contextual nature of AI and of its ethical concerns is also highlighted in the abovementioned Ethics Guidelines for Trustworthy AI.

[34] See e.g. European Commission, 'Artificial Intelligence for Europe', 25 April 2018, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51625; High-Level Expert Group on AI, 'Ethics Guidelines for Trustworthy AI'; High-Level Expert Group on AI, 'Policy and Investment Recommendations for Trustworthy AI', 26 June 2019.

which is driven by a 'good cause' yet tends to blind those in charge of the technology's adoption for its risks, also propels AI's ubiquity and hence gives cause to the concerns that this paper deals with.

When examining the features of the algorithmized world, it can first be noted that AI is a general-purpose technology.[35] The algorithms that compose AI systems can be rendered operable in any domain, and the same type of AI system can also be repurposed for different contexts. Consider, for instance, the case of a Japanese AI-enabled computer vision system designed to distinguish between different types of pastries, to automate the check-out process at the register of a cafeteria-style shop. This system was later repurposed to help distinguish between different types of cancer cells.[36]

Second, the systems' automated nature means they can carry out their algorithmic processes incessantly, without any human intervention (after their initial programming) and without requiring a break to eat or sleep (as long as their energy supply is ensured). This enables organizations to provide their services 24/7 if they wish so. In addition, the fact that AI systems can process vast amounts of data in a short amount of time, also makes them amenable to increase the efficiency of processes. Besides leading to time-savings, these efficiencies can also help reduce costs.[37]

Third, precisely because they are able to peruse a high quantity of data in a short amount of time, AI systems can at times help improve the accuracy of human decision-making.[38] A well-known example concerns the fact that, for certain types of cancer, AI systems can more accurately detect the disease than radiologists based on the analysis of CT scans.[39] Similar

---

35 This feature often leads to its comparison with electricity or oil. See e.g. European Commission, 'Artificial Intelligence for Europe'. However, see also: Samm Sacks and Justin Sherman, 'Calling Data "the New Oil" Could Hurt Efforts to Protect Privacy', *Slate Magazine* (blog), 13 June 2019, https://slate.com/technology/2019/06/data-not-new-oil-kai-fu-lee-china-artificial-intelligence.html.

36 James Somers, 'The Pastry A.I. That Learned to Fight Cancer', *The New Yorker*, 18 March 2021, https://www.newyorker.com/tech/annals-of-technology/the-pastry-ai-that-learned-to-fight-cancer.

37 McKinsey, 'Notes from the AI Frontier: Modeling the Impact of AI on the World Economy', Discussion Paper (McKinsey, September 2018).

38 Evidently, given the typically probabilistic nature of data-driven AI systems, a margin of error remains, and the potential reduction of human error does not eradicate the risk of machine-error.

39 Elizabeth Svoboda, 'Artificial Intelligence Is Improving the Detection of Lung Cancer', *Nature* 587, no. 7834 (18 November 2020): S20–22. Note also the qualification made in the article about the limitations of these AI systems: *"… humans are better at learning quickly about the minutiae of unusual lung cancer cases. AI systems, on the other hand, excel at flagging common types of early cancerous lesion, having been trained on data sets that include thousands of such cases.*" The technology is thus meant to complement and augment the work of radiologists rather than replacing them.

methods have been deployed for the detection of other diseases[40], but can also be found in different domains, from manufacturing quality control[41] to agricultural applications.[42]

Fourth, AI systems are often described as able to take over tasks that are unduly repetitive and hence intellectually unstimulating for humans, or tasks that are dangerous and would unduly expose humans to risks. Examples of the former are AI systems that assist with simple administrative acts or repetitive manual labor, while the latter can concern AI systems that help clean a nuclear site or assist in mine-explorations. Reports that highlight the advantages of such AI systems typically stress the fact that this frees up valuable time that workers can spend on safer or more interesting tasks, rather than replacing their jobs.[43]

Fifth, AI systems also enable the personalization of services and products at lower cost, thereby reconciling the scale of mass consumption with the need for individual tailoring. Examples concern the personalization of medical treatments based on an analysis of patients' specific traits and how this correlates with the way other patients reacted, or the personalization of online advertisement based on the individual preferences of consumers as inferred from their online behavior. A similar technique is also used to personalize people's newsfeeds on social media. The upside of this news-personalization is that we no longer need to peruse through the overload of information we are confronted with, since the algorithm can learn our preferences and do this for us. As we shall see, there is also a downside to this, from the potential creation of echo chambers[44] to mass-surveillance.[45]

The above advantages – and the economic benefits they can generate – not only led to the increasingly widespread adoption of AI, but also triggered a global 'race to AI'.[46] Countries

---

40 See for instance Chenyu Shi et al., 'Use of Convolutional Neural Networks for the Detection of U-Serrated Patterns in Direct Immunofluorescence Images to Facilitate the Diagnosis of Epidermolysis Bullosa Acquisita', *The American Journal of Pathology*, 28 June 2021.

41 Hamidey Rostami, Jean-Yves Dantan, and Lazhar Homri, 'Review of Data Mining Applications for Quality Assessment in Manufacturing Industry: Support Vector Machines', *International Journal of Metrology and Quality Engineering* 6, no. 4 (2015); Han Ding et al., 'State of AI-Based Monitoring in Smart Manufacturing and Introduction to Focused Section', *IEEE/ASME Transactions on Mechatronics* 25, no. 5 (2020): 2143–54.

42 Guan Wang, Yu Sun, and Jianxin Wang, 'Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning', *Computational Intelligence and Neuroscience* 2017 (5 July 2017): e2917536; Andreas Kamilaris and Francesc X. Prenafeta-Boldú, 'Deep Learning in Agriculture: A Survey', *Computers and Electronics in Agriculture* 147 (1 April 2018): 70–90.

43 Evidently, whether or not AI takes over someone's job does not depend on any AI system, but on the human employer with decision-making power over such matters. As will be stressed further below, this responsibility extends not only to the actions of AI systems, but also to decisions regarding the development and deployment of those systems. Leaving in the middle what the effects of the widespread use of AI on the job market will be, contrary to the language often used in media, AI systems never 'steal' jobs or 'create' jobs – human beings do. See for instance the linguistic shortcuts in this regard in: Priya Mohanty, 'Do You Fear Artificial Intelligence Will Take Your Job?', Forbes, 6 July 2018, https://www.forbes.com/sites/theyec/2018/07/06/do-you-fear-artificial-intelligence-will-take-your-job/. For a critique on this phenomenon, see Arleen Salles, Kathinka Evers, and Michele Farisco, 'Anthropomorphism in AI', *AJOB Neuroscience* 11, no. 2 (2 April 2020): 88–95; Smuha, 'Laten We Intelligenter Zijn Wanneer We Het over Artificiële Intelligentie Hebben'.

44 Justin D Martin et al., 'From Echo Chambers to "Idea Chambers": Concurrent Online Interactions with Similar and Dissimilar Others', *International Communication Gazette*, 16 February 2021.

45 Karen Yeung, 'Five Fears about Mass Predictive Personalization in an Age of Surveillance Capitalism', *International Data Privacy Law* 8, no. 3 (1 August 2018): 258–69.

46 Yuval Noah Harari, 'Who Will Win the Race for AI?', *Foreign Policy* (blog), accessed 15 July 2020, https://foreignpolicy.com/gt-essay/who-will-win-the-race-for-ai-united-states-china-data/.

across the world promulgated AI strategies, aiming to be 'world leader in AI'.[47] AI is, in fact, often perceived as a technology that can help progress towards an ever-better human condition, a view that bears strong affinities with progressivism – an idea that will be revisited later.

At this stage, a number of caveats must be made. First, the abovementioned benefits do not automatically materialize. The strengths of AI are limited to narrowly defined tasks, and depend on how they are designed and which data they use. Second, even when benefits are achieved, this does not mean they actually benefit *all*. Oftentimes, those who already find themselves in a beneficial position are best-placed to reap those benefits, whereas those who are in a vulnerable position are not necessarily better off. In addition, given the hype surrounding this technology, the capacities of AI applications are sometimes oversold[48] and hyperbolic statements about the benefits that certain systems can generate – followed by disappointing results – are not uncommon.[49] Finally, it is not because an AI system is developed or deployed with good intentions, that it is also developed and deployed in a good manner, with due attention to its unintended consequences. Let us therefore move towards a closer examination of what the risks of AI entail.

3.2 The bad

AI systems do not operate in a vacuum. They are always embedded in a broader environment, which is composed not only of the direct surrounding of the machine itself, but also of the broader network of individuals, organizations, cultures, languages, laws and customs. In other words, AI systems are 'socio-technical' systems, as they have an influence on – and are influenced by – their social environment.[50] The mutual influencing process between AI and society[51] renders it indispensable to consider the societal paradigm under which the algorithmization of the world, and the global race to do so, is enabled.

---

47 Smuha, 'From a "Race to AI" to a "Race to AI Regulation"', 58.

48 Kate Crawford also describes the phenomenon of so-called 'Potemkin AI', whereby a product is sold as an autonomous system for marketing purposes, but in fact primarily relies on human labor behind the scenes, often in very dire labor circumstances. See Crawford, *Atlas of AI*, 65.

49 A recent example is the use of AI to help counter the COVID-19 pandemic. Soon after COVID-19 broke out, numerous tech developers enthusiastically started designing and deploying AI systems with great expectations about the way these could be used against the virus. However, the results were disappointing and AI was not able to deliver its promise. See Will Douglas Heaven, 'Hundreds of AI Tools Have Been Built to Catch Covid. None of Them Helped.', MIT Technology Review, 30 July 2021, https://www.technologyreview.com/2021/07/30/1030329/machine-learning-ai-failed-covid-hospital-diagnosis-pandemic/.

50 High-Level Expert Group on AI, 'Ethics Guidelines for Trustworthy AI', 8 April 2019; Gordon Baxter and Ian Sommerville, 'Socio-Technical Systems: From Design Methods to Systems Engineering', *Interacting with Computers* 23, no. 1 (2011); Andreas Theodorou and Virginia Dignum, 'Towards Ethical and Socio-Legal Governance in AI', *Nature Machine Intelligence* 2, no. 1 (2020): 10–12; Shakir Mohamed, Marie-Therese Png, and William Isaac, 'Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence', *Philosophy & Technology*, 12 July 2020; Pekka Ala-Pietilä and Nathalie A. Smuha, 'A Framework for Global Cooperation on Artificial Intelligence and Its Governance', in *Reflections on Artificial Intelligence for Humanity*, ed. Bertrand Braunschweig and Malik Ghallab (Cham: Springer International Publishing, 2021), 237–65; Gry Hasselbalch, *Data Ethics of Power* (Edward Elgar Publishing, 2021).

51 Consider in this regard also the seminal papers of Langdon Winner and Melvin Kranzberg respectively, discussing the societal impact of technology more generally – equally applicable to the sphere of AI: Langdon

This paradigm is characterized by a stark belief that the more data we collect, the better human decision-making will be. In other words: to make the world a better place, we need big data, based on which we can take the 'best' course of action. The idea behind this paradigm is simple. Human beings have inherent cognitive limitations.[52] Even if they read and study all their lives, they will never be able to process all that is out there. Moreover, whichever action they take may be tainted by human error, sleep deprivation or cognitive biases. AI systems, however, are not deemed to suffer from those limitations. While they may be unable to undertake a variety of intelligent tasks at once[53], their ability to analyze vast amounts of data renders them not just *able* but – according to this paradigm – also *better placed* to decide on the 'best' course of action.

Considering the above, there are three assumptions that underpin the paradigm of the algorithmized world. The first concerns the assumption that if we 'let the data speak',[54] we will be able to identify the 'best' decision from a range of options (a). The second concerns the assumption that the cognitive biases of human decision-makers are not present with non-human machines – and that those machines are, instead, able to generate 'truly objective' outcomes (b). The third assumption pertains to the notion that all human decisions can be broken down to data points: elements that can be quantified, measured or otherwise (digitally) collected to enable their computation (c). Unfortunately, each of these assumptions is wrong. In what follows, I explain why, and draw attention to a number of risks arising from this fallacy.

(a) Why we cannot just 'let the data speak'

First, the idea that we can delegate responsibility for difficult decisions to algorithms if only we feed them with sufficient data, rests on a misunderstanding of both the nature of human problems

Winner, 'Do Artifacts Have Politics?', *Daedalus* 109, no. 1 (1980): 17; Melvin Kranzberg, 'Technology and History: "Kranzberg's Laws"', *Bulletin of Science, Technology & Society* 15, no. 1 (1 February 1995): 5–13.

52 Thomas L. Griffiths, 'Understanding Human Intelligence through Human Limitations', *Trends in Cognitive Sciences* 24, no. 11 (1 November 2020): 873–83.

53 AI systems today are defined as so-called 'narrow AI' as opposed to 'general AI'. Narrow AI is programmed to carry out a specific task in a specific domain, such as diagnosing a particular type of cancer or winning a game of Alpha Go. While it can be highly intelligent in carrying out its task and in some situations even able to surpass human performance, it is unable to perform functions outside its programmed scope. Thus, an AI system programmed to win Alpha Go, even if defeating the best human player in the world, will not be able to recommend a movie or take out the dog for a walk. While this does not mean that narrow AI cannot produce results that are unexpected by its developers (e.g. through the misalignment of values in the optimization function of the system), its capacities and limitations entirely rest upon the humans that programmed it. Contrary to narrow AI, general AI is characterized by its ability to autonomously carry out a multitude of complex tasks across various domains, including a level of moral sentience that renders it an independent agent. Today, and in any foreseeable future, no such AI exists (even if scientists across the world are actively working towards its creation and attracting significant funding for this endeavor). The focus of this paper exclusively concerns narrow AI.

54 The approach to deduce the model based on the gathered data rather than structuring the data around a pre-defined model is one of the landmarks of the modern approaches to AI-based data analytics, such as machine learning techniques. Already in 1973, famous French mathematician Jean-Paul Benzécri introduced the idea of "letting the data speak for themselves", stressing that "*Le modèle doit suivre les données, non l'inverse"*, in J.-P. Benzécri, *L'analyse des données*. *2: L'analyse des correspondances*, Leçons sur l'analyse factorielle et la reconnaissance des formes et travaux du Laboratoire de statistique de l'Université de Paris VI (Paris: Dunod, 1973). See also François Husson, Julie Josse, and Gilbert Saporta, 'Jan de Leeuw and the French School of Data Analysis', *Journal of Statistical Software* 73, no. 6 (2016): 16.

and the nature of data. Two domains need to be distinguished here: the positive and the normative, which are often confused in the context of AI. Let me elucidate this with an example.

Consider an AI system that assists a law firm in hiring a new lawyer. It is one thing for the system to help determine whether a candidate obtained the required degree for the job. The system is not asked to determine what *should* be a prerequisite for a candidate, but merely to assess whether the – already human-defined prerequisite – is attained. It is, however, another thing for the system to help determine, for instance based on data of past candidates, which qualities render someone 'right' for the job. The first task belongs to the positive realm. A human (in this case, lawmakers) already decided that the possession of a certain degree is needed to be 'right' for the job, and the algorithm is merely deployed to peruse the candidate's data to determine whether this is the case. The normative decision (the fact that a degree is warranted to be considered for the job) is transparent, and precedes the positive one. The second task, however, belongs to the normative realm. The law firm here outsources the normative decision of what makes a candidate 'right' for a job to the system – which will in turn rely on the codified optimization function and the data it was fed by human beings. The normative decision may not be transparent here, because we do not know which factors were flagged as normatively relevant, yet it is there.

When humans seek to understand the best approach to deal with a problem, they already – often implicitly – have an idea of what the ideal outcome would be, based on their values and preferences. While data analysis can help determine what the best course of action might be given a value X, it will never be able to determine the value that humans should strive for *as such*. Thinking otherwise is a conflation of the normative and the positive.[55] No matter how attractive this fallacy might be portrayed, it is a naïve approach at best and leads to a dangerous discharge of responsibility for normative decision-making at worst. Those at power may prefer to believe or explain that their decisions are not actual decisions at all, since complex data models perhaps even unanimously pointed towards the same course of action. Yet the question is of course: what were those models optimized for? Ultimately, the optimization decision remains a human one, even if the step of that decision remains hidden behind layers of code.

A note should be made here regarding this 'hidden' aspect of AI systems, which is sometimes referred to as a 'black-box' problem, denoting AI's opacity and inscrutability.[56] The opacity of AI systems can increase the challenge of identifying and addressing their problems – yet it should also be nuanced. Not all opacity that surrounds AI has something to do with the 'black-box' problem. Opacity in AI can manifest itself in at least three – non-exclusive – ways, which concern: the fact that AI is used, the way in which AI is used and the way in which AI works. Only the third type of opacity relates to AI's black-box issue.

---

[55] Also known as falling into the trap of the is/ought fallacy. The articulation of the is-ought problem is most notably ascribed to David Hume. See David Hume, *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects and Dialogues Concerning Natural Religion*, ed. L. A. Selby-Bigge (Oxford: Clarendon Press (1896), 1739). See also Max Black, 'The Gap Between "Is" and "Should"', *The Philosophical Review* 73, no. 2 (1964): 165–81.

[56] See e.g., Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press, 2015).

First, there can be opacity around the deployment of an AI system. Given AI's digital and frictionless nature, individuals may not always know that an AI system is being used to make recommendations or decisions about them.[57] In addition, since AI systems can mimic human behavior – for instance as a chatbot – they can also be used to deliberately pretend they are a human being. This lack of transparency around the use of AI can adversely affect our privacy, as well as our human autonomy and dignity.[58]

Second, there can be opacity around the way in which an AI system is used. This concerns, for instance, the type of data fed into the system, what the system was optimized for, what the underlying assumptions are that the system is built on, and how the outcomes of the system are used within the organization that deploys the system. These value-laden choices are rarely made transparent, which strengthens the idea that those choices do not exist. The idea is, of course, mistaken, yet risks overshadowing the fact that the developers of these systems, who are usually already in a position of power, can retain their power precisely through the non-contestable configuration of these systems.

Third, there can be opacity around the AI system's inner workings – and this is where the black-box problem actually comes in. For some data-driven systems (for instance, those based on deep learning), it cannot be explained how their internal decision-making processes work. This renders it difficult to evaluate whether these processes are based on robust and fair methods. However, even when such inscrutable systems are used, transparency regarding the two points above – which *is* possible – can already enhance the system's oversight. Yet the 'black-box' concept arising with this third type of opacity is an easy pretense to also bring the *human-chosen* opacity of the two other points under this notion, and avoid external scrutiny.[59]

(b) Why AI systems are not free from bias

Second, the myth that AI systems generate objective outcomes can likewise be busted. Although AI systems lack moral agency and are not inherently motivated by (self-) interests or prejudices, they reflect the prejudices and cognitive biases of their developers. Human biases can also manifest themselves through the data fed into the system.[60] If a facial recognition system is only trained on datasets showing pictures of white men, the system will not be able to recognize

---

57 Think of remote facial recognition systems that might scan our faces without us being aware of this, but also of AI-enabled online psychographic targeting that may be used to manipulate us into buying certain products or believing certain dis- or misinformation. See also Kate Crawford, *Atlas of AI* (New Haven: Yale University Press, 2021), 109.

58 Catelijne Muller, 'The Impact of Artificial Intelligence on Human Rights, Democracy and the Rule of Law', Report Prepared in the Context of the Council of Europe's Ad Hoc Committee on AI (CAHAI) (Strasbourg: Council of Europe, 24 June 2020), https://www.coe.int/en/web/artificial-intelligence/cahai.

59 Crawford, *Atlas of AI*, 12.

60 Harini Suresh, 'The Problem with "Biased Data"', Medium, 26 April 2019, https://medium.com/@harinisuresh/the-problem-with-biased-data-5700005e514c; Frederik J. Zuiderveen Borgesius, 'Strengthening Legal Protection against Discrimination by Algorithms and Artificial Intelligence', *The International Journal of Human Rights*, 25 March 2020, 1–22; Eirini Ntoutsi et al., 'Bias in Data-Driven Artificial Intelligence Systems—An Introductory Survey', *WIREs Data Mining and Knowledge Discovery* 10, no. 3 (2020).

women and people of color as accurately.[61] Thus, the outcomes of AI systems – and the biases they reflect – hinge on the human decisions laying at their basis. The fact that a broad range of societal domains, and hence also the data collected from these domains, are still plagued by inequalities and discriminations, renders the unchecked use of AI systems liable to perpetuate and even exacerbate unjust biases – at scale. A well-known example concerns the AI-system used by Amazon to assist the company's recruitment decisions. [62] Based on the data from employee performances in the past, the system was trained to assess which new candidates would be best suited. However, as this dataset primarily contained information about white male employees – in part due to the fact that, historically, the company counted many more male than female employees – the system evaluated CVs from female candidates more negatively.

In the meantime, ever more data is being collected and processed, often also including 'personal data'.[63] Given the sensitivity of such data – it reveals so much about individuals that it can be considered as part of their persona, much in the same way as a limb[64] – in Europe, the right to personal data protection was elevated to a fundamental right.[65] Evidently, this right is under increased tension with data-driven AI systems, which can be highly privacy-intrusive – whether deliberately or not. The personalization of products and services for individuals, for instance, relies on the ability to collect sufficient data about them to profile them, and to subsequently make inferences and predictions about their character or behavior.[66] Furthermore, the

---

61 Buolamwini and Gebru, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification'; Timnit Gebru, 'Race and Gender', in *The Oxford Handbook of Ethics of AI*, by Timnit Gebru, ed. Markus D. Dubber, Frank Pasquale, and Sunit Das (Oxford University Press, 2020), 251–69.

62 Jeffrey Dastin, 'Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women', *Reuters*, 10 October 2018, sec. Retail, https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G. One might respond by saying that the solution would be to simply exclude information about a candidate's gender from the dataset, yet the problem is that other data – which in first instance do not concern a person's gender – might nevertheless reveal one's gender through correlation. The example of Apple's credit card, launched in 2019 and almost immediately criticized for offering women less credit than men, is very telling in this regard. The algorithm underneath it was explicitly designed to be 'blind' for gender, based on which its designers claimed that it could impossibly be biased. Nevertheless, other proxies that correlated with gender still lead to gender discrimination, which was less difficult to detect precisely because the system was 'gender-blinded'. See also, e.g., Frederik J. Zuiderveen Borgesius, 'Discrimination, Artificial Intelligence, and Algorithmic Decision-Making' (Strasbourg: Council of Europe - Directorate General of Democracy, 2018), 13. Will Knight, 'The Apple Card Didn't "See" Gender—and That's the Problem', *Wired*, 2019, https://www.wired.com/story/the-apple-card-didnt-see-genderand-thats-the-problem/.

63 Under EU General Data Protection Regulation (GDPR), such information is broadly defined as "*any information that relates to an identified or identifiable living individual*". See European Parliament and Council, 'Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)', OJ L 119 (2016).

64 Luciano Floridi, 'On Human Dignity as a Foundation for the Right to Privacy', *Philosophy & Technology* 29, no. 4 (December 2016): 307–12.

65 Orla Lynskey, *The Foundations of EU Data Protection Law* (Oxford: Oxford University Press, 2015).

66 Such predictions not only rely on the data of the individual that is being assessed, but also on the data of many other individuals, and how their traits correlate. On AI-enabled profiling, see e.g., Mireille Hildebrandt and Bert-Jaap Koops, 'The Challenges of Ambient Law and Legal Protection in the Profiling Era', *Modern Law Review* 73, no. 3 (2010): 428–60; Stefanie Hänold, 'Profiling and Automated Decision-Making: Legal Implications and Shortcomings', in *Robotics, AI and the Future of Law*, ed. Marcelo Corrales, Mark Fenwick, and Nikolaus Forgó, Perspectives in Law, Business and Innovation (Singapore: Springer, 2018), 123–53. See also Salomé Viljoen, 'Democratic Data: A Relational Theory for Data Governance', *Available at SSRN: Https://Ssrn.Com/Abstract=3727562*, November 2020.

combination of different data-sets might yield new possibilities for analysis, thus incentivizing organizations not only to collect more data, but also to keep it stored in case an opportunity arises to use it in another context.[67] Consequently, the amplified incentives to gather data from individuals can eventually also lead to *de facto* mass surveillance.[68]

(c) Why not everything that matters can be measured

Third, the above postulations rely on the ability to measure and quantify phenomena, which can then be translated into digital and analyzable data. The idea of quantifiability is also closely related with controllability: once we understand certain occurrences, we can use this knowledge to control and shape them to our benefit. While this approach has helped us make significant scientific advances, the transposition thereof from the scientific to the social realm is less smooth. As is well-known, complex social phenomena are not readily translatable to quantifiable metrics, and hence require the mediation of indicators and proxies.[69] Since there is, however, hardly ever a one-on-one relationship between the social phenomenon and the indicator, something gets lost.[70] This deficit is precisely what AI systems likewise suffer from, as they are utterly dependent on (the quality of) these indicators and proxies.[71]

Consider the example of an AI system used by a bank to evaluate someone's 'creditworthiness'. While a person's 'creditworthiness' is difficult to quantify, there are several elements that could provide an indication thereof, such as the money a person has on her account, potential debts she might carry, or her income. These elements are more easily quantifiable and could thus be used as a proxy for 'creditworthiness', which is what the bank is ultimately after. Crucially, however, the proxies chosen are not always reflective of the sought-after phenomenon: this will depend on the soundness of the assumptions made by the system developer.[72] Moreover, similar approaches are taken to map a person's 'character', or a person's propensity to 'act criminally'

---

67 For a commercial actor, these insights could focus on the way in which a certain service or product can best be commercially marketed based on individuals' preferences. For a law enforcer, these insights could focus on the physical places in which most crimes occur, and where police resources should hence be prioritized. See also Crawford, *Atlas of AI*, 95.

68 Yeung, 'Five Fears about Mass Predictive Personalization in an Age of Surveillance Capitalism'.

69 Sally Engle Merry, 'Measuring the World: Indicators, Human Rights, and Global Governance', *Current Anthropology* 52, no. S3 (April 2011): S83–95; Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Houghton Mifflin Harcourt, 2013).

70 See also Geoffrey C. Bowker and Susan Leigh Star, 'Building Information Infrastructures for Social Worlds — The Role of Classifications and Standards', in *Community Computing and Support Systems: Social Interaction in Networked Communities*, ed. Toru Ishida, Lecture Notes in Computer Science (Berlin, Heidelberg: Springer, 1998), 231–48; Luke Stark, 'Algorithmic Psychometrics and the Scalable Subject', *Social Studies of Science* 48, no. 2 (April 2018): 204–31.

71 Rachel Thomas and David Uminsky, 'The Problem with Metrics Is a Fundamental Problem for AI', *Ethics of Data Science Conference 2020*, 19 February 2020, http://arxiv.org/abs/2002.08512.

72 Furthermore, certain proxies may be relevant in theory, but are illegal to take into account in practice given that they can lead to unjust discrimination. A hypothetical study might indicate, for instance, that over the past 50 years women were overall less creditworthy than men (without necessarily explaining the historical reasons for this). While, on that basis, banks could choose to make the assumption that sex is a valid indicator of someone's creditworthiness, they are in principle not allowed to take this factor into account in their evaluation, since sex is a prohibited discrimination ground. Note that I am careful in using the word 'sex' rather than 'gender' here, given that 'gender' is currently not included in the list of prohibited discrimination grounds of the Charter of Fundamental Rights of the European Union ("the EU Charter") or the European Convention on Human Rights ("the ECHR").

– all of which are not readily translatable to AI-analyzable data. Yet the human condition cannot simply be reduced to mathematical utility functions.[73] This reductionist approach – taking a detached and impersonal perspective – to human action, risks ignoring essential aspects of our humanity, for the sake of efficiency.[74]

In sum, while the intentions behind the wide-spread adoption of AI systems might be (at least, to some extent) noble, the paradigm of the algorithmized world is underpinned by assumptions that stand on shaky ground. As long as we hold on to them, we risk erroneously confusing normative choices with positive statements, we risk overlooking the biased nature of AI systems, and we risk forgetting that not everything that matters can be quantified. This can lead to the escape of responsibility for normative decisions, the perpetuation and exacerbation of discrimination, and over-reliance on indicators that do not reflect reality.

It is important to stress, at this stage, that each of the abovementioned risks can manifest itself not only due to conscious action – for instance, the deliberate choice to deploy AI in a discriminatory way – but also due to negligent inaction. Like all tools, AI systems can purposely be used to cause harm.[75] Yet what interests me here, is the harm that can be caused as a side-effect of AI's use, without a malicious purpose.[76] The absence of due diligence by the systems' developers and deployers to examine and justify the assumptions that underlie their system, is of considerable concern. Whether this absence stems from a lack of awareness of the problem, or from technological hubris and a disinterest in questioning one's methods, does not alter the fact that the system can negatively affect those subjected thereto.

3.3 The ugly

In the chapters above, I analyzed the algorithmized world by considering the 'good' – namely the benefits that AI can offer such world – as well as the 'bad' – namely the problematic paradigm underpinning it and the risks associated therewith. There is, however, a further step we must take to get a more comprehensive view of what is at stake. We have, thus far, primarily considered the impact of AI in an insulated fashion. I described how an AI system that operates based on flawed assumptions can lead to erroneous and discriminatory outcomes, and how an AI system that collects data from individuals can impact their right to privacy. It is, however, necessary to extend our gaze beyond individual AI systems, and to consider their scale as well.

---

73 Foreword by Danielle Allen, x, in Hannah Arendt, *The Human Condition* (University of Chicago Press (2019), 1958).

74 This also holds as regards our subjective temporal existence. See in this regard Michael L. Morgan, *The Cambridge Introduction to Emmanuel Levinas* (Cambridge: Cambridge University Press, 2011), 162.

75 Miles Brundage and et al., 'The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation', February 2018, https://maliciousaireport.com/.

76 Some might argue that the advancement of private interests by large corporations in a capitalist society – in particular through AI-enabled surveillance tools – is 'malicious' in and of itself, yet for the purpose of this paper, I will treat this as the desire to entrench (market) power rather than the desire to cause deliberate harm. For a discussion of 'surveillance capitalism' and the role AI systems play in this regard, see e.g. Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, 1st ed. (New York: PublicAffairs, 2019).

Today, AI systems are no longer used in a limited number of delineated domains. The algorithmized world is characterized by the wide-spread permeation of AI, and by our ever-increased reliance thereon for important decisions. Rather than operating in an isolated manner, AI systems are part of a broader network of systems running through our entire societal infrastructure – physically as well as digitally. Accordingly, the risks flagged above are not isolated outliers, but they are present at scale due to their cumulative impact. It is precisely this scale that renders AI systems liable to not just affect individual rights, but to uproar our core values, and to shake the normative foundations of society.[77] This shaking process may be a slow one – it follows the pace of the normalization of AI's ubiquity – yet it is also a profound one.

Elsewhere, I made a distinction between individual, collective and societal harms raised by AI, which can help clarify what is at stake.[78] Individual harm occurs when one or more interests of an individual are wrongfully thwarted.[79] This is the case, for instance, when an AI system operates based on incorrect assumptions or biased data sets. Consider the example of a biased facial recognition system used by law enforcement to identify criminals, which disproportionally misidentifies people of color leading to their wrongful arrest – and hence to their discrimination.[80] Of course, the thwarting of such an interest does not occur in isolation from a social, historical and political context.[81] Nevertheless, in this scenario, at the receiving end of the harm stands an identifiable individual who, due to her skin color, is being discriminated.

Collective harm occurs when one or more interests of a collective or group of individuals are wrongfully thwarted. Just as a collective consists of the sum of individuals, so does this harm consist of the sum of harms suffered by individual members of the collective. The use of the abovementioned biased facial recognition system can, for instance, give rise to collective harm, where it thwarts the interest of a collective of people – namely the people of color who are subjected to the AI system – not to be discriminated against.[82] The collective dimension hence arises from the accumulation of similarly thwarted individual interests.[83]

Societal harm occurs when one or more interests of society are wrongfully thwarted. It is hence not concerned with the interests of a particular individual or collective of individuals, but instead focuses on harm to an interest held by society at large, going over and above the sum of

---

77 Karen Yeung, 'Responsibility and AI - A Study of the Implications of Advanced Digital Technologies (Including AI Systems) for the Concept of Responsibility within a Human Rights Framework' (Council of Europe, DGI(2019)05, September 2019).

78 Nathalie A. Smuha, 'Beyond the Individual: Governing AI's Societal Harm', *Internet Policy Review*, 2021.

79 Joel Feinberg, 'Harm to Others', in *The Moral Limits of the Criminal Law - Volume 1: Harm to Others* (New York: Oxford University Press, 1984).

80 This example is, unfortunately, not hypothetical. See e.g. Kashmir Hill, 'Wrongfully Accused by an Algorithm', *The New York Times*, 24 June 2020, sec. Technology, https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html.

81 Thomas W. Simon, *Democracy and Social Injustice: Law, Politics, and Philosophy* (Rowman & Littlefield, 1995).

82 Crawford, for instance, makes this point as regards collective or group privacy. Crawford, *Atlas of AI*, 111. See also Linnet Taylor, Luciano Floridi, and Bart van der Sloot, eds., *Group Privacy: New Challenges of Data Technologies* (Cham: Springer International Publishing, 2017).

83 Andrew Kernohan, 'Accumulative Harms and the Interpretation of the Harm Principle', *Social Theory and Practice* 19, no. 1 (1993): 51–72.

individual interests. This can be clarified by revisiting the previous example. We established that, by making use of such a biased system and wrongfully thwarting the interest of an individual of color, the system's deployer can cause individual harm. The accumulation of the harm done to individuals of color at the collective level, entails collective harm. Yet a third type of harm is at play. Whether individuals are colored or not, and whether they are subjected to the particular AI system or not, they share an interest to live in a society that does not discriminate against people based on their skin color and that treats people equally. That interest is different from the interest not to be discriminated against, and can hence be distinguished from the individual or collective harm done to those directly subjected to the AI system. Societal harm can hence be assessed as a *sui generis* type of harm.[84]

Each of these three types of harm is, of course, problematic. Yet distinguishing between them – and in particular, conceptualizing societal harm – can help us understand the more fundamental impact that the wide-spread use of AI systems can have in the algorithmized world. Concretely, it clarifies that every member of society – regardless of whether she is directly subjected to a particular AI system or not – can be adversely affected thereby, through the way in which the systemic and problematic characteristics of the system are impacting the values that are commonly shared in modern liberal democracies. Beyond equality, there are numerous other societal interests that can be affected by AI systems[85], such as the interest in democracy[86] and the rule of law.[87] Consider, for instance, the way in which AI systems can be used to collect and analyze personal data for political profiling purposes, and to subsequently subject individuals to tailored manipulation techniques – this being the downside of AI's ability to enable personalization.[88]

Scandals like Facebook/Cambridge Analytica demonstrated that AI-enabled psychographic targeting can be used as a tool to try shaping political opinions with the specific purpose of influencing election outcomes.[89] Similar techniques can also be used to subliminally drive people towards commercial products and services (if I profiled you as prone to gambling, I can target my gambling advertisement more effectively towards you), polarizing or extreme standpoints (if I profiled you as prone to appreciate racist content, I can target such content more

---

84 Inspiration for treating societal harm as a sui generis type of harm can be found in Durkheim's sui generis treatment of 'society'. See Emile Durkheim, *L'éducation Morale* (Paris: Alcan, 1925).

85 Another pertinent example concerns the impact of the use of AI systems on the environment, and in particular their massive ecological footprint. See e.g. Emma Strubell, Ananya Ganesh, and Andrew McCallum, 'Energy and Policy Considerations for Deep Learning in NLP', 5 June 2019, http://arxiv.org/abs/1906.02243; Karen Hao, 'Training a Single AI Model Can Emit as Much Carbon as Five Cars in Their Lifetimes', *MIT Technology Review* (blog), 6 June 2019, https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/; Crawford, *Atlas of AI*.

86 Brkan, 'Artificial Intelligence and Democracy'.

87 Roger Brownsword, 'Technological Management and the Rule of Law', *Law, Innovation and Technology* 8, no. 1 (2 January 2016): 100–140; Bayamlıoğlu and Leenes, 'The "Rule of Law" Implications of Data-Driven Decision-Making'; Hildebrandt, 'Algorithmic Regulation and the Rule of Law'.

88 Frederik J. Zuiderveen Borgesius et al., 'Online Political Microtargeting: Promises and Threats for Democracy', *Utrecht Law Review* 14, no. 1 (9 February 2018): 82; Brkan, 'Artificial Intelligence and Democracy'.

89 Jim Isaak and Mina J. Hanna, 'User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection', *Computer* 51, no. 8 (August 2018): 56–59.

effectively towards you) or misinformation (if I profiled you as prone to believe in conspiracy theories, I can target those messages more effectively towards you). Given that we continuously share more data about ourselves – whether as consumers or as citizens – the inferences that can be made about us steadily increase. Consequently, the manipulative practices that can be exacerbated by AI systems can take place at an ever-wider scale and can lead to increasingly effective results. The potential harm associated with these practices – from election interference to polarization – goes beyond the persons that are directly manipulated, but affects society as a whole, hence constituting societal harm.

If we now recall the issues highlighted above – and in particular the problematic assumptions that underpin the paradigm of the algorithmized world – it becomes more evident that the operation of those systems at scale, in combination with the opacity surrounding them, can have a fundamental impact on our being. We increasingly delegate human decisions to machines, all the while maintaining (1) the unjustified idea that these machines merely execute positive tasks where instead normative choices are being delegated, together with a delegation of the responsibility for those choices; (2) the mistaken belief that – inevitably partial – datasets and optimization functions can lead to 'objective' outcomes; and (3) the overreliance on proxies and indicators to capture complex social phenomena that are impossibly reducible to quantifiable metrics. To these issues, we can add the aforementioned increase in (personal) data collection and the enabling of mass surveillance, the entrustment of authority over human-impacting decisions to fallible AI systems, and the search to influence human decisions through AI-enabled manipulation. All of these concerns, naturally, bring us very far from the homeliness of our intersubjective existence, where human dignity, privacy, equality, democracy, and respect for the other's individuality and autonomy are treasured.

The question is then: what does this mean for our being-in-the-world? How should we make sense of these tendencies which seem to alter core tenets of our existence, and what action should we take to deal with their adverse effects? It is at this point that ethics is supposed to come to the rescue, by analyzing as well as providing a critique of, and answer to, the challenges raised above. But is it also delivering?

4. WHY CURRENT ETHICS DISCOURSE FALLS SHORT OF ITS PURPOSE

The algorithmization process that steadily took place over the past decades has been accompanied by a parallel, though more recent, development: increased attention for the ethics of AI. Beyond questions of theoretical philosophy[90] – for instance, about the meaning of

[90] Next to Turing's aforementioned paper, consider, for instance, Patricia Smith Churchland, *Neurophilosophy: Toward a Unified Science of the Mind-Brain*, 2nd print (Cambridge (Mass.): MIT press, 1986); Raymond Kurzweil, *The Age of Spiritual Machines: When Computers Exceed Human Intelligence* (New York: Viking, 1999); Luciano Floridi, 'The Ontological Interpretation of Informational Privacy', *Ethics and Information Technology* 7, no. 4 (December 2005): 185–200; Aziz Zambak and Roger Vergauwen, 'Artificial Intelligence and Agentive Cognition: A Logico-Linguistic', *Logique et Analyse* 52, no. 205 (2009): 57–96; Roger Vergauwen, 'Will Science and Consciousness Ever Meat? Complexity, Symmetry and Qualia', *Symmetry* 2, no. 3 (September 2010): 1250–69; Mark Coeckelbergh, 'When Machines Talk: A Brief Analysis of Some Relations between Technology and Language', *Technology and Language* 1, no. 1 (2020): 22–27.

concepts like 'artificial' and 'intelligence' – the adoption of AI in the very practical contexts of our everyday lives also started giving rise to questions of practical philosophy or ethics.[91]

Critique on the implementation of new technologies is not new. Consider, for instance, the nineteenth-century Luddite movement, which originated amidst British textile workers who challenged the introduction of automated looms in textile factories – and even set out to destroy these machines.[92] Their concerns were primarily of socio-economic nature, since they feared that their carefully learned craft – and hence their jobs – had become redundant. While not constituting the main focus of this paper, many of those fears are also present today due to the advent of AI.[93] Another example, likewise dating back to the 19th century, concerns the use of the then-modern technology of 'instantaneous photography', which – coupled with the widespread circulation of newspapers – raised new possibilities to intrude on people's private lives, eventually leading to the conceptualization of a right to privacy.[94]

As described above, the advent of AI systems – whether considered on a case-by-case basis or as a broader phenomenon – is not devoid of risks. While not all of these risks are new,[95] they manifest themselves in novel ways in light of the specific characteristics of AI, and are increasingly under scrutiny, thanks to the work of committed researchers, vocal practitioners, active civil society organizations and engaged journalists.[96] Some AI-applications also led to a public outrage when the harm they caused became more widely known,[97] which in turn triggered

---

91 Mark Coeckelbergh, *AI Ethics*, The MIT Press Essential Knowledge Series (Cambridge, Mass: MIT Press, 2020). Within the domain of ethics, it is the field of applied ethics that deals with the question of what a person ought to do in a specific situation or domain of action. The 'ethics of AI' can hence be seen as a sub-field of applied ethics that focuses on the ethical conundrums raised by the development and use of AI systems. See also High-Level Expert Group on AI, 'Ethics Guidelines for Trustworthy AI', 9.

92 Steven E. Jones, *Against Technology: From the Luddites to Neo-Luddism*, 1st ed. (New York: Routledge, 2006).

93 See e.g. Mohanty, 'Council Post'; European Commission, 'Artificial Intelligence for Europe'.

94 Samuel D. Warren and Louis D. Brandeis, 'The Right to Privacy', *Harvard Law Review* 4, no. 5 (1890): 193–220.

95 See e.g. Smuha, 'Beyond a Human Rights-Based Approach to AI Governance'.

96 See e.g. Kate Crawford and Meredith Whittaker, 'The AI Now Report - The Social and Economic Implications of Artificial Intelligence Technologies in the Near Term' (New York: The AI Now Institute, 2016), https://ainowinstitute.org/AI_Now_2016_Report.pdf; O'Neil, *Weapons of Math Destruction*; Buolamwini and Gebru, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification'; Karen Yeung, 'Responsibility and AI - A Study of the Implications of Advanced Digital Technologies (Including AI Systems) for the Concept of Responsibility within a Human Rights Framework' (Council of Europe, DGI(2019)05, September 2019); Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code*, 1 edition (Medford, MA: Polity, 2019); Knight, 'The Apple Card Didn't "See" Gender—and That's the Problem'; Carole Cadwalladr, 'Fresh Cambridge Analytica Leak "Shows Global Manipulation Is out of Control"', *The Guardian*, 4 January 2020, sec. UK news, http://www.theguardian.com/uk-news/2020/jan/04/cambridge-analytica-data-leak-global-election-manipulation; AlgorithmWatch, 'Automating Society Report 2020', October 2020, https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/10/Automating-Society-Report-2020.pdf; Karen Hao, 'He Got Facebook Hooked on AI. Now He Can't Fix Its Misinformation Addiction', MIT Technology Review, 11 March 2021, https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/.

97 Isaak and Hanna, 'User Data Privacy'; Human Rights Watch, 'China's Algorithms of Repression: Reverse Engineering a Xinjiang Police Mass Surveillance App' (Human Rights Watch, 1 May 2019), https://www.hrw.org/report/2019/05/01/chinas-algorithms-repression/reverse-engineering-xinjiang-police-mass-surveillance; Sean Coughlan, 'Why Did the A-Level Algorithm Say No?', *BBC News*, 14 August 2020, sec. Family & Education, https://www.bbc.com/news/education-53787203; Will Bedingfield, 'Everything

a broader discussion on AI's ethical concerns. As a consequence, a growing number of actors[98] have sought to map and analyze these concerns, and to reflect on how to address them. As however noted in the introduction, most of these reflections take for granted the technology's ubiquity, and focus on how ethics can be brought into the technology rather than the other way around – which may lead to an incomplete comprehension of the nature of the problem and to an inadequate response. In what follows, I provide a brief overview of current ethics discourse in the context of AI (4.1) and explain why it seems to fall short of its purpose (4.2). I then propose a different approach by seeking an Archimedean point – grounded in human intersubjectivity rather than technology – to pave the way for the next steps of our inquiry (4.3).

4.1 An overview of current 'AI ethics' discourse

The level of attention given to AI ethics has exponentially increased over the past few years, and by now arguably reached almost the same heights as attention to AI itself – which is not an easy feat. The boom of academic articles across the globe dealing with this subject is not the only testimony to this development.[99] The contemporary approach to AI ethics manifests itself through a variety of initiatives, which I briefly describe in this section. Importantly, the thread running through all those initiatives is their aim to harness the beneficial potential of AI systems, all the while minimizing their risks by ensuring their trustworthiness.[100]

For instance, besides classical education programs, a number of MOOCs (massive open online courses) were developed to draw attention to AI's ethical risks, so as to spread knowledge and awareness about them and stimulate AI developers to take them into account.[101] These educational initiatives are also accompanied by a plethora of policy documents – originating from a wide range of organizations – that set out what the ethical risks of AI systems are, and what precautions can be taken to avoid them.[102] In addition, various companies that design or

---

That Went Wrong with the Botched A-Levels Algorithm', *Wired UK*, 19 August 2020, https://www.wired.co.uk/article/alevel-exam-algorithm.

98 These include not only ethicists but also governments, international organizations, non-governmental organizations, companies, public institutions, legal scholars and others.

99 Note also the creation of new academic journals devoted entirely to this subject, e.g., the recently established *AI and Ethics* published by Springer (2020, eds. John MacIntyre and Larry Medsker), and the *AI Ethics Journal* published by AIRES (2019, ed. Aaron Hui).

100 Although, in theory, both of these aims are typically put forward together, in practice, depending on the initiative-takers and their stance towards the paradigm of the algorithmized world, the former often appears to take precedence over the latter. See e.g. Hagendorff, 'The Ethics of AI Ethics'.

101 Consider, in this regard, for instance, the 'Elements of AI' course, originating in Finland and translated in all EU languages through European Commission funding. While this AI course has a component on AI and ethics, the course developers also prepared a dedicated spin-off course entirely focused on the ethics of AI: https://ethics-of-ai.mooc.fi/. See also the MOOC developed by Agoria on Sustainable AI in Business, likewise freely accessible: https://www.agoria.be/sustainable-ai-in-business/en/.

102 High-Level Expert Group on AI, 'Policy and Investment Recommendations for Trustworthy AI'; Council of Europe Ad Hoc Committee on Artificial Intelligence (CAHAI), 'Feasibility Study'; OECD, 'Recommendation of the Council on Artificial Intelligence'; Information Commissioner's Office, 'Guidance on AI and Data Protection' (ICO, July 2021), https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/guidance-on-ai-and-data-protection/; UK Department for Education, 'Realising the Potential of Technology in Education: A Strategy for Education Providers and the Technology Industry', 2019, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/791931/DfE-Education_Technology_Strategy.pdf.

deploy AI systems started developing technical tools – both for internal use and for use by their customers and other interested organizations – to verify whether a given AI system is biased[103], or to enhance its privacy-friendliness, for instance by working with decentralized data processing methods such as federated learning.[104]

In addition to these technical tools to 'make AI more ethical', a rising number of consultancies are now also offering their services to evaluate the AI systems used by organizations and to verify how their 'ethical' nature can be increased.[105] This development can only be described as one in which ethics is increasingly seen as a product or service. Such view is even explicitly promoted in certain papers as a pragmatic approach to implement ethics into the development and deployment processes of AI, coined as 'ethics-as-a-service'.[106] A similar pragmatism can also be found in the adoption of 'AI ethics principles' that companies promise to commit to.[107] As part of that commitment, companies sometimes also set up internal or external AI ethics advisory boards, to guide their use of AI in line with ethical principles.[108]

An important development within this approach, concerns the promulgation of AI ethics guidelines and checklists by national, supranational and international organizations, typically intended as self-help tools for AI developers and deployers. One of the most well-known examples thereof concerns the abovementioned *Ethics Guidelines for Trustworthy AI*,[109] drafted by the European Commission's High-Level Expert Group on Artificial Intelligence. This group consisted of a range of stakeholders that were brought together to advise the European Commission on AI policies in the European Union, and to prepare a set of practical ethics guidelines.[110] These Guidelines set out, inter alia, seven essential requirements that AI systems should meet throughout their entire lifecycle to be considered 'Trustworthy'.[111] To operationalize these requirements, the Guidelines also contain an assessment list with detailed questions that guide AI developers and deployers through the questions they should ask themselves to enhance their system's trustworthiness. Such guidelines are, however, non-binding. AI developers and deployers can thus freely choose whether or not to respect them –

---

103 Consider, for instance, IBM's AI Fairness 360 tool. IBM Research, 'Introducing AI Fairness 360, A Step Towards Trusted AI', September 2018, https://www.ibm.com/blogs/research/2018/09/ai-fairness-360/.

104 See e.g. Huadi Zheng, Haibo Hu, and Ziyang Han, 'Preserving User Privacy for Machine Learning: Local Differential Privacy or Federated Machine Learning?', *IEEE Intelligent Systems* 35, no. 4 (2020): 5–14.

105 See, for instance, Deloitte's ethical AI programme. Deloitte, 'Bringing transparency and ethics in AI', Deloitte Netherlands, accessed 12 August 2021, https://www2.deloitte.com/nl/nl/pages/innovatie/artikelen/bringing-transparency-and-ethics-into-ai.html.

106 Jessica Morley et al., 'Ethics as a Service: A Pragmatic Operationalisation of AI Ethics', *Minds and Machines* 31, no. 2 (1 June 2021): 239–56.

107 Consider, for instance, Google's AI ethics principles: Google, 'AI at Google: Our Principles', Google AI, accessed 12 August 2021, https://ai.google/principles/. Yet consider also: Khari Johnson, 'AI Ethics Pioneer's Exit from Google Involved Research into Risks and Inequality in Large Language Models', VentureBeat, 3 December 2020. https://venturebeat.com/2020/12/03/ai-ethics-pioneers-exit-from-google-involved-research-into-risks-and-inequality-in-large-language-models/.

108 Note that these developments are primarily driven by large tech companies that have the budget to do so.

109 High-Level Expert Group on AI, 'Ethics Guidelines for Trustworthy AI'.

110 See also Smuha, 'The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence'.

111 The seven requirements for Trustworthy AI concern: respect for human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental well-being; and accountability.

something that has been heavily criticized, given the extent of AI's risks and the wide-scale harm it can cause.[112]

As a consequence – in what can be considered as the culmination of current developments in AI ethics – attempts are now made by regulators to translate these guidelines into binding legislation. The most prominent actor in this regard is the European Commission, which published a proposal for an AI Regulation[113] in April 2021 that largely codifies the ethics requirements proposed by the High-Level Expert Group on AI in its Guidelines. While it will likely still take months if not years before the proposal is adopted by the European Parliament and Council, and while there is much scope for improvement,[114] this step represents a clear commitment on behalf of the EU to tackle AI's adverse impact on the health, safety and fundamental rights of individuals.[115] A similar approach is currently undertaken by the Council of Europe, which aims to develop a legal instrument to protect human rights, democracy and the rule of law against AI's risks.[116] However – bearing in mind the profoundness of AI's impact for the human condition, which will be further elaborated on in Chapter 5 – these initiatives risk being woefully insufficient to help us make sense of, and tackle, the extent of AI's (adverse) consequences.

4.2 The limits of 'ethics-as-a-service'

Before explaining why they fall short, let me stress that I consider each of the above initiatives to be welcome developments. Working in a complementary way, they are able to spread substantial knowledge of AI-related concerns, and to provide an important layer of protection against many of its excesses, including in a legally enforceable manner. I am hence not putting in question these initiatives' *necessity*. Instead, I am questioning their *sufficiency*. Considering the scale of the problems discussed, these initiatives – even cumulatively – seem to 'mop the floor while the tap is open'[117] rather than providing a fundamental critique of, and response to, the profound challenges of the algorithmized world.

---

112 See for instance Michael Veale, 'A Critical Take on the Policy Recommendations of the EU High-Level Expert Group on Artificial Intelligence', *European Journal of Risk Regulation*, 23 January 2020, 1–10; Hagendorff, 'The Ethics of AI Ethics'.

113 European Commission, Proposal for a Regulation of the European Parliament and the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts.

114 Nathalie A. Smuha et al., 'How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act' (Social Science Research Network, 5 August 2021), https://papers.ssrn.com/abstract=3899991.

115 It should be noted that the Council of Europe, which consists of 47 member states, is also working on a binding legal instrument aimed at safeguarding human rights, democracy and the rule of law against AI's adverse effects – possibly in the form of an international convention. See e.g. Council of Europe Ad Hoc Committee on Artificial Intelligence (CAHAI), 'Feasibility Study'.

116 See e.g. Council of Europe Ad Hoc Committee on Artificial Intelligence - CAHAI, 'Feasibility Study' (Strasbourg: Council of Europe, 17 December 2020). Building on the CAHAI's work, the Council of Europe's Committee of Ministers has now given its successor, the Committee on Artificial Intelligence (CAI), the mandate to draft said legal instrument.

117 There seems to be no direct translation of this Flemish proverb in English, but I trust that the reader will get the gist.

The role of ethics is limited to providing a quick assessment of – and ideally, a 'quick fix' for – the adverse impact of AI on individuals, including their right to safety, their right to non-discrimination and their right to privacy. Questions that go beyond the impact of individual AI systems on individual human beings, are largely left out of scope.[118] In particular, the societal harm raised by AI that I conceptualized above as going beyond individual and collective harm, seems to be difficult to grasp within this discourse. In other words: these initiatives might counter some of the 'bad', but they fail to counter most of 'the ugly'. Moreover, they do not question the societal paradigm that enables the algorithmized world in the first place.

It should at this stage be noted that current ethics discourse in the context of AI has not been left entirely uncriticized. Concerns have been raised about the instrumentalization of AI ethics by commercial actors in particular, with the purported aim of keeping stricter regulation at bay – a phenomenon denoted as ethics-washing[119], analogous to the concept of green-washing.[120] Some critics even imply that ethics discourse should be left aside in the context of AI since it is 'toothless', and that the conversation of AI's risks should be managed by legal discourse instead.[121] Differently than ethical standards and the advice of ethics boards, legal rules are binding and can be enforced. Now that regulators are starting to translate ethical standards into legal instruments, the focus of this strand of criticism has shifted to the – lack of – comprehensiveness of these laws.

While the legal rules that are being developed are certainly not flawless, and while certain actors may undoubtedly seek to instrumentalize AI ethics discourse for their own purposes, I believe the above line of critique is deficient in two ways: it conflates the role of ethics and law, and it still does not ensure that the fundamental issues raised by the algorithmized world are assessed. Let me clarify both. First, calling out the limited scope of AI ethics as a poor – or deliberate – substitute for legal rules reduces ethics to a "*softer version of the law*",[122] thereby showing a misunderstanding of the respective and complementary role of each of these domains.[123] Under

---

118 Nathalie A. Smuha, 'Beyond the Individual: Governing AI's Societal Harm', *Internet Policy Review*, 2021. See also Daniel Greene, Anna Lauren Hoffmann, and Luke Stark, 'Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning', *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.

119 See, for instance, Wagner, 'Ethics As An Escape From Regulation. From "Ethics-Washing" To Ethics-Shopping?'; Brent Mittelstadt, 'Principles Alone Cannot Guarantee Ethical AI', *Nature Machine Intelligence* 1, no. 11 (November 2019): 501–7; Jobin, Ienca, and Vayena, 'The Global Landscape of AI Ethics Guidelines'; Rodrigo Ochigame, 'The Invention of "Ethical AI": How Big Tech Manipulates Academia to Avoid Regulation', *The Intercept* (blog), 20 December 2019, https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/.

120 William S. Laufer, 'Social Accountability and Corporate Greenwashing', *Journal of Business Ethics* 43, no. 3 (2003): 253–61; Kent Walker and Fang Wan, 'The Harm of Symbolic Actions and Green-Washing: Corporate Actions and Communications on Environmental Performance and Their Financial Implications', *Journal of Business Ethics* 109, no. 2 (2012): 227–42; Klarissa Lueg and Rainer Lueg, 'Detecting Green-Washing or Substantial Organizational Communication: A Model for Testing Two-Way Interaction Between Risk and Sustainability Reporting', *Sustainability* 12, no. 6 (January 2020): 2520.

121 This deficiency is also pointed out in Bietti, 'From Ethics Washing to Ethics Bashing'; Anaïs Rességuier and Rowena Rodrigues, 'AI Ethics Should Not Remain Toothless! A Call to Bring Back the Teeth of Ethics', *Big Data & Society* 7, no. 2 (1 July 2020).

122 See Rességuier and Rodrigues, 'AI Ethics Should Not Remain Toothless! A Call to Bring Back the Teeth of Ethics', who rely on a quote of Jobin, Ienca, and Vayena, 'The Global Landscape of AI Ethics Guidelines'.

123 Smuha, 'The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence'.

this critique, ethics is no longer seen as a rigorous mode of inquiry that requires the theorization and justification of one's moral stance, but as an inefficient form of justice, or worse, "*a form of cover-up or façade for unethical behavior*."[124] Second, no matter how rigorously ethics guidelines are translated to legal instruments, and no matter how strictly they are enforced, such instruments will be equally ill equipped to provide a more profound critique of how the widespread use of AI is impacting our mode of existence, not only at the individual but also at the societal level. As long as this underlying impact is not problematized, the proposed solutions – even if legally binding in nature – will offer only limited solace.

As I indicated in this paper's introduction, an explanation of the deficiency of current ethics discourse can be found in the fact that – critical as it may try to be – it remains stuck within the overarching technological paradigm that governs the algorithmized world. As part of this paradigm, ethics is meant to help us orient AI systems towards their best possible use, and to mitigate the issues they pose so that we can enjoy their benefits to the fullest. The ubiquity of AI systems, and our extensive reliance on technology more generally, is considered as a given. Furthermore, the three problematic assumptions that underpin the algorithmized world's paradigm are not fundamentally put in question, nor are the broader implications thereof for our way of being. The starting point is the facticity of technology, as well as the unstoppable progress it brings forth. The harms of AI are seen as isolated instances which, in the overall picture, do not beg foundational questions but merely need to be avoided to continue pursuing the progress that AI promises to deliver. While this view may be a good fit for a progress-oriented narrative of the world, in which history is a rationally intelligible and linear process, the narrative's validity can at the very least be seriously questioned.

Consider in this regard Walter Benjamin's *anti*-philosophy of history, which cautions us for the "*dangerous political complacency which follows an uncritical belief in the inevitability of progress in human affairs*".[125] The image of the Angelus Novus, a 1920 monoprint by Paul Klee which Benjamin acquired in 1921, provides the visual representation of his argumentation that historical progress – or the idea that we are generally moving into an ever better future – is an illusion.[126] In his '*Theses on the Philosophy of History'*, Benjamin describes the painting as follows:

> A Klee painting named Angelus Novus shows an angel looking as though he is about to move away from something he is fixedly contemplating. His eyes are staring, his mouth is open, his wings are spread. This is how one pictures the angel of history. His face is turned toward the past. Where we perceive a chain of events, he sees one single catastrophe which keeps piling wreckage upon wreckage and hurls it in front of his feet. The angel would like to stay, awaken the dead, and make whole what has been smashed. But a storm is blowing from Paradise; it has got caught in his wings with such violence that the angel can no longer close them. The storm irresistibly propels him into the future

[124] Bietti calls this approach 'ethics-bashing'. See Bietti, 'From Ethics Washing to Ethics Bashing'.

[125] James Connelly, 'Facing the Past: Walter Benjamin's Antitheses', *The European Legacy* 9, no. 3 (June 2004): 319.

[126] See Walter Benjamin, *Theses on the Philosophy of History* (New York: Schocken Books, 1968). See also Ronald Beiner, 'Walter Benjamin's Philosophy of History', *Political Theory* 12, no. 3 (1984): 423–34.

> to which his back is turned, while the pile of debris before him grows skyward. This storm is what we call progress.[127]

The passage clarifies that, for as much as we may be looking for it, there is no linear chain of events in history. The angel gazes at the past and has his back to the future – as the future is unknown to us – and is facing a storm that propels him ever further into the future amidst a growing pile of debris.[128] At the same time, the flight backwards into the future cannot be stopped, and "*the destructiveness of linear empty time, pushes the angel even as it inhibits him*."[129] The point which Benjamin makes – and which echoes in the writings of other critical theorists, from Adorno[130] to Allen[131] – forces us to approach any uncritical progress narrative with care, no matter how powerful the technology that can allegedly propel such progress. Importantly, considering historical materialism as an illusion does not entail a denial of the fact that technology significantly contributed to human well-being. But it does entail a rejection of the narrative which often accompanies technological advances – and AI in particular – which claims they are part of an inevitable, linear and continuous idea of 'Progress', while barely allowing for criticism regarding its societal impact.

Given the above, the role of today's ethics discourse as part of this technological paradigm of progress risks being reduced to either a legitimization of AI's ubiquity ("it's everywhere, but if we layer ethics-as-a-service on it, we can rubber-stamp it"), or the melioration thereof ("it's everywhere, and if we layer ethics-as-a-service on it, we can make it even better") – neither of which is satisfactory. The question is then, how can we lift ethics from this limited discourse and deploy it to formulate a more fundamental critique? The answer lays in the question: we need to transcend the paradigm of the algorithmized world in which this discourse is still too deeply embedded. Rather than making a utilitarian cost-benefit analysis between AI's benefits and risks, or even abandoning the hope we lay on ethics altogether[132], we should reinstate ethics' more fundamental role by turning the tables around and seeking an Archimedean point[133] outside the technology. This does not mean a denial of AI's ubiquity or a call for its elimination. Nor does it mean finding a synthesis between the various 'thesis and anti-thesis' tensions that

---

127 Benjamin, *Theses on the Philosophy of History*.

128 As noted by Susan Handelman, "*this storm also seems to represent the destructive aspects of a revolution, whose purgation alone can bring any 'progress' to the ruins of history*". See Susan Handelman, 'Walter Benjamin and the Angel of History', *CrossCurrents* 41, no. 3 (1991): 346.

129 See Handelman, 348. As Handelman points out, while Benjamin's words paint a rather grim image, Gershom Scholem, a Jewish philosopher and close friend of Benjamin offers a glimmer of hope. He allows us to link Benjamin's description of the 'angel of history' to the Talmudic messianic thinking, in which catastrophe and redemption are entwined. According to a well-known Talmudic legend, the blackest day in Jewish history – namely the day of the catastrophic destruction of the Temple, which initiated the Jewish exile – was also the day that the Messiah was born. See Gershom Scholem, *On Jews and Judaism in Crisis: Selected Essays*, ed. Werner J. Dannhauser (New York: Schocken Books, 1976).

130 Theodor W. Adorno, *Negative Dialektik. Jargon der Eigentlichkeit* (Frankfurt am Main: Suhrkamp, 1973).

131 Amy Allen, *The End of Progress: Decolonizing the Normative Foundations of Critical Theory* (New York: Columbia University Press, 2016).

132 This solution has been espoused by numerous ethics critics in the past. See Morgan, *The Cambridge Introduction to Emmanuel Levinas*, 5.

133 See also footnote 6.

AI poses to our mode of being.[134] Instead, we can take these tensions seriously[135] by seeking to interpret them through a meta-technological discourse that puts ethics first.[136]

4.3 Intersubjectivity as Archimedean point for a meta-technological discourse

It is precisely here, when adopting an Archimedean point to reflect upon AI's impact on the human condition, that I believe Jewish authors can offer a valuable contribution. As noted in the introduction, drawing inspiration from such authors, I propose to seek this point in intersubjective relationality. Arguably, the attention of Jewish thinkers to intersubjectivity can be linked – directly or indirectly – to the centrality thereof in Judaism more generally. This can be illustrated by, for instance, examining one of the most important texts reflecting Jewish ethics, namely *Pirkei Avot*.[137] As part of the Mishna, the text constitutes a compilation of moral teachings stemming from Rabbinic Jewish tradition.[138] Rather than being concerned with ritual and legal practices as many Jewish texts from that period are, Pirkei Avot "*is a work that consists purely of timeless life wisdom*",[139] thus rendering it quite unique in Jewish ethical literature. Interestingly, the very first *mishnah* or teaching that this work starts with, immediately draws the reader's attention to the primacy of ethics and interhuman relationships, which comes even *before* the human relationship with God. Let us consider it more closely:

> Moses received the Torah from Sinai and transmitted it to Joshua; and Joshua to the Elders; and the Elders to the Prophets; and the Prophets transmitted it to the Men of the Great Assembly. They said three things: Be deliberate in judgment; develop many students; and make a fence for the Torah.[140]

A few remarks can be made. First, this *mishna* firmly establishes the authority of the writers of the *Pirkei Avot* by tracing their link to Moses, to whom God revealed the Torah. Second, its tri-partite advice to be deliberate in judgment (work on oneself), develop many students (help others) and make a fence for the Torah (keep up the Jewish tradition) not only reflects the Jewish ethos, but also indicates attention to ternary thinking and the triangular relationship between the

---

134 Consider the parallel reasoning regarding the ethical concerns arising from globalization in Anckaert, 'Globalisation and the Tragedy of Ethics', drawing also on Adorno's 'negative dialectic' which pushes us to retain a critical reflection rather than seeking a synthesis that risks drowning protest voices. See Adorno, *Negative Dialektik. Jargon der Eigentlichkeit*.

135 Anckaert, *God, Wereld en Mens*, 13.

136 The aim of this paper is to put ethics first, both literally and figuratively. Literally because we are starting our inquiry from ethics rather than from technology. Figuratively, this already foreshadows my reliance on intersubjectivity as Archimedean point, which is also essential for Emanuel Levinas when he establishes that 'ethics is first philosophy'. See in this regard also Morgan, *The Cambridge Introduction to Emmanuel Levinas*, 3.

137 *Pirkei Avot* is often translated as *The Ethics of the Fathers*, in light of its ethical content, though its literal translation is *The Chapters of the Fathers.*

138 *Pirkei Avot* is part of the *Mishna*, the first written version of the Jewish oral tradition (sometimes referred to as the 'The Oral Torah' as opposed to 'The Written Torah' on which the oral tradition forms a comment), redacted around the start of the 3rd century. It contains 63 volumes or tractates which discuss all domains of Jewish law, as well as the tractate 'Avot' ('Fathers') that deals with Jewish ethics. See also George Robinson, *Essential Judaism: A Complete Guide to Beliefs, Customs & Rituals* (New York: Atria Books, 2016).

139 Shmuly Yanklowitz, *Pirkei Avot: A Social Justice Commentary* (New York: CCAR Press, 2018), xii.

140 Pirkei Avot, Chapter 1:1, translation in Yanklowitz, *Pirkei Avot: A Social Justice Commentary*.

self, other and Torah. Third, and most important for our purpose, is the fact that this *mishna* does not focus on God's handing over of the Torah to Moses, or on the relationship between God and humans. Instead, by referring to the fact that Moses 'receives' the Torah from 'Sinai', the focus lays on Moses' transmission of the Torah to Joshua, and on the subsequent chain of inter-human transmissions. As explained by Dr. Yanklowitz:

> By beginning in this manner, Mishna 1:1 describes the Torah's primary focus on human relationships. Were this Mishnah to focus on God first, then ethics – which are matters between human beings – would necessarily be considered second. Ethics become the foundation for a covenantal relationship with the Divine. The Sages impart this message from the start. The entire Torah enterprise requires relationship.[141]

This emphasis on interhuman relationships, and hence on the intersubjective nature of our being, is a vital hallmark of Jewish tradition.[142] It also runs as a thread through twentieth-century Jewish philosophy, during which the horrors of the wars only enhanced attention to ethics. Yanklowitz, for instance, directly links the aforementioned Mishnaic primacy of ethics to the work of Emmanuel Levinas.[143] About seventeen centuries after the Mishna's publication, Levinas posited "*ethics as first philosophy*"[144], thereby giving it precedence over ontology. Levinas grounds the primacy of ethics for human existence in the face-to-face encounter between the self and the other.[145] We may be trying to make things known to us by drawing on concepts that we are familiar with, yet are inevitably faced with an 'Other' that is different from us, imposes its dependence on us – which Levinas also links to the other's suffering – and puts us in a position of responsibility.[146] The primacy of ethics, which Levinas advocates, can be considered as a deliberate move against Heidegger's prioritization of the *Dasein*,[147] which structures reality around Being and hence remains averse to plurality.[148] More generally, this move can also be seen as a broader critique of Western philosophy's inwards-turning approach, focusing primarily on the subject rather than on relationality and the role of the other.[149]

A similar criticism against Heidegger – and Western philosophy – was formulated by Hannah Arendt[150], who likewise emphasized intersubjectivity throughout her work. As noted by Anya Topolski, differently than Levinas, she conceptualizes this approach through the notion of *plurality* rather than *alterity*.[151] Like Levinas' understanding of 'Other' as alterity, Arendt

---

141 Yanklowitz, 3.
142 See also Robinson, *Essential Judaism*.
143 Yanklowitz, *Pirkei Avot: A Social Justice Commentary*, 4.
144 Emmanuel Levinas, *Totalité et Infini - Essai Sur l'exteriorité* (Paris: Le Livre de Poche (2021), 1971). See also Steven Crowell, 'Why Is Ethics First Philosophy? Levinas in Phenomenological Context', *European Journal of Philosophy* 23, no. 3 (2015): 564–88.
145 Michael L. Morgan, *Discovering Levinas* (Cambridge: Cambridge University Press, 2007), 64.
146 Levinas, *Totalité et Infini - Essai Sur l'exteriorité*.
147 Anya Topolski, *Arendt, Levinas and a Politics of Relationality*, Reframing the Boundaries: Thinking the Political (Lanham: Rowman and Littlefield, 2015), 17.
148 Topolski, 20.
149 Morgan, *The Cambridge Introduction to Emmanuel Levinas*.
150 See for instance Arendt's critique on the loss of common sense. Arendt, *The Human Condition*.
151 For a comparison and, in particular, a bringing into dialogue of the works of Levinas and Arendt – who did not seem to have had a dialogue with each other in real life, despite being born in the same year and having frequented similar circles – see Topolski, *Arendt, Levinas and a Politics of Relationality*.

cherishes the irreducible uniqueness of human beings, and rejects any view through which plurality is eliminated in favor of a unity – which hampers individual freedom and risks leading towards totalitarianism. Instead, meaning can be found in intersubjective relationships, which necessitates a 'common world' in which a diversity of human beings can act together.[152] Arendt considers that the presence of other human beings is a prerequisite to life, and an essential precondition to experience meaning.[153]

A generation earlier, Franz Rosenzweig – whose work *The Star of Redemption*[154] heavily influenced the writings of numerous twentieth-century Jewish thinkers – already broke ground by rejecting Hegelian idealism's lack of attention to the particular, in favor of a dialogical approach.[155] Rosenzweig proposed a renewal of thinking (the 'new thinking'), in which the idealist ontological *monism* – by which the entire reality is explained through thinking – is ruptured by an ontological *pluralism*, marked by relationality.[156] At first instance, this concerns the ternary relationality between *God*, the *world* and the *human*, each of which stands in relation to but is irreducible to the other, and relies upon this relationality to exist.[157] Yet the metaphysical individuality of these entities eventually also leads to the – need to respect the – relational yet 'in-dividual' nature of human beings, and leave space to speak with and hear the Other. One can also recall the work of a friend of Rosenzweig, Martin Buber, whose short yet influential book *I and Thou* plainly posits that "*Man becomes an I through a You*".[158] Buber, like Rosenzweig, posits a ternary structure, consisting of an *I*, *Thou* and *It*, and cautions for the distortion of human relationships when *Thou* or, more modernly, *You*, is treated as an *It*.[159]

In sum, the aforementioned approaches all reject an excessive focus on the subject *qua* subject, or on totalizing unifying ideas, and rebalance this focus towards one that includes the relationship between I and the Other, and intersubjectivity. It is this focus that I will maintain when considering what the algorithmized world means for the human condition.

---

152 "*Men in so far as they live and move and act in this world, can experience meaningfulness only because they can talk with and make sense to each other and to themselves*", in Arendt, *The Human Condition*, 188; Topolski, *Arendt, Levinas and a Politics of Relationality*, 45.

153 Arendt, *The Human Condition*, 4.

154 Franz Rosenzweig, *Der Stern Der Erlösung*, trans. Alexandre Derczanski and Jean-Louis Schlegel (Paris: Editions du Seuil (2003), 1976).

155 Ibid. See also Anckaert, *God, Wereld en Mens*; Luc Anckaert, 'Language, Ethics, and the Other between Athens and Jerusalem. A Comparative Study of Plato and Rosenzweig', *Philosophy East and West* 45, no. 4 (1 January 1995): 545–67.

156 See also Anckaert, 'Franz Rosenzweigs Stern der Erlösung. Een hermeneutische en retorische benadering', in which this is described as follows: "*Rosenzweig's critique of idealism can be summarized in a few points: the particular is incorporated into the universal; behind the appearing reality a real reality is postulated; reality is equated with reason. To this end, a unitary principle (one-dimensionality, analogy, emanation) is used as an instrument*" (my translation).

157 Rosenzweig, *Der Stern Der Erlösung*. Anckaert, *God, Wereld en Mens*.

158 Martin Buber, *I and Thou*, trans. Walter Kaufman (New York: Simon and Schuster (2000), 1923), 80.

159 See in this regard also the foreword of Walter Kaufman in Buber, *I and Thou*.

5. AI'S IMPACT ON THE HUMAN CONDITION

The human condition is a vast subject. I will hence delineate my analysis by focusing on three aspects thereof in particular: the way we think or *rationality* (5.1), the way we engage with each other or *alterity* (5.2), and the way we experience time or *history* (5.3). At the outset of this analysis, it is important to keep in mind that these three aspects are entwined, and that they can be perceived as standing in a ternary relationship with each other. The way we think about the world and about ourselves, and the way we rationalize the choices we make, for a large part also depends on the role we assign to the other, and the way we engage with others. It also depends on the manner in which we perceive the passing of time and deal with our history, as well as the way we create and identify our place – and the place of others – in the past, present and future.[160] Our experience of time, furthermore, occurs in a context with other subjects, rather than as isolated individuals.[161] Accordingly, the insights that will be formulated under the three sections below should not be considered separately, but rather as complementary to – and even reinforcing – each other.

5.1 Rationality – Algorithms and Binarity

Human rationality, or the way we reason and think, constitutes a core aspect of our being. Against an intersubjective backdrop, in which our being in the world is a being-with-others, this rationality is qualified by an ethos that reflects the inherently pluralistic human existence. For Levinas, we are, in fact, fundamentally *ethical* beings – rather than for instance fundamentally *rational*[162], or driven by desires or emotions.[163] While Levinas does not speak here in normative terms, but describes a self-grounded ontological reality, this has certain implications regarding not only our actual but also our desired form of reasoning. An important corollary of our plural existence concerns the rejection of totalizing systems thinking – a type of thinking that Rosenzweig, Levinas and Arendt all argued against, each in their own terms.

Talking about the risks she sees for the human condition in the modern age – including the risks associated with the search for artificial life and wide-spread automation – Hannah Arendt states that: "*What I propose, therefore, is very simple: it is nothing more than to think what we are doing*."[164] Although the *Human Condition* focuses particularly on the *Vita Activa* rather than the *Vita Contemplativa*, her book is riddled with attention for the ways in which the particular

---

160 See e.g. Rosenzweig, *Der Stern Der Erlösung*.

161 Consider for instance Emmanuel Levinas, *Le Temps et l'Autre*, 11th ed. (Paris: Presses Universitaires de France (2014), 1979).

162 A distinction can be made between two uses of the term 'rationality': first, it can be used to denote the way in which we think more generally; but, second, it can also be used to denote a stricter, scientific way of thinking or approaching the world which not only prioritizes reason, but also seeks a detachment from social and emotional factors to take an 'objective' perspective (technocratic rationality). See e.g. Niklas Andreas Andersen, 'The Technocratic Rationality of Governance - the Case of the Danish Employment Services', *Critical Policy Studies*, 28 December 2020, 1–19. In this section, I make the case that the algorithmic rationality – which is infused by the second approach – is different from, and stands in tension with, our way of thinking more generally when considering an intersubjective outlook.

163 Morgan, *The Cambridge Introduction to Emmanuel Levinas*, 4.

164 Arendt, *The Human Condition*, 5.

rationality of modernity counters the ideal of human action, which for Arendt is par excellence an activity that humans do together and is hence of primary value.

In this section, let us hence take up Arendt's proposal and think what we are doing in the algorithmized world, by comparing the intersubjective rationality with the algorithmized one. In what follows, I draw particular attention to the shift towards an approach grounded in binary thinking, whereby things risk getting lost in translation due to a reductionist view of human beings and social phenomena (a). I also examine the risk of eliminating spontaneous opportunities for goodness in favor of a systematizing and totalizing idea of the Good (b). Finally, I assess the erosion of the role of speech – which is an essential element of our rationality – and how this erosion correlates with the risk of dehumanization (c).

(a) From plurality to binarity

Our complex world constitutes a web of people, objects, events and ideas, which stand in relation to each other and can be considered from a virtually infinite number of perspectives. Although we try to make sense of these things by structuring them through language, concepts and rules, these are mostly social constructs that, necessarily, only capture part of the rich reality. My mother is not just a 'mother', but also a 'wife', a 'daughter', a 'colleague', a 'friend', a 'consumer', a 'reader', a 'Belgian', a 'woman' and much more. In addition, all individuals have an inherent dignity and merit being treated with respect for their own multifaceted individuality.[165] Depending on the particular context, one or more of these aspects of an individual can be considered over and above others – for instance when people are categorized or classified for a given purpose – even if that aspect is but one part of a more comprehensive picture. It is thus possible, yet always partial, to place people and things into a category – even if we do this on a daily basis.[166] Consider, for instance, the categorization of behaviors that are 'legal' and 'illegal', or distinctions between people that are 'single', 'married' or 'divorced'.[167] These conventional concepts, limited as they may be to reflect the richness of reality, are how we structure our world, and also help us to express our thoughts to one another. Yet as long as we find ourselves in an intersubjective environment, we can draw attention to the limitations of these structures and explicate why a certain categorization is erroneous and requires correction, or we can provide nuance, or we can ask for additional options or categories given that we share the concept's *meaning*. Furthermore, the inherently linguistic nature of these concepts also means they are, in principle, always open for interpretation, contestation and adaptation.

The way data categorization and analysis take place, has been accelerated exponentially by the capacity to capture data in digitized form, at scale, and unleash computations on it with a speed that surpasses that of the human mind. Under traditional AI systems, the categories in which data are classified are typically still codified by human programmers, who thereby decide (and

---

165 This does not mean that the individual should be prioritized over a community at all costs, but it does mean that – even when considering a community – it should not be lost out of sight that it consists of a plurality of individuals who, albeit standing in relation to each other, have their own beliefs, thoughts and life projects.

166 See e.g., Bowker and Star, 'Building Information Infrastructures for Social Worlds — The Role of Classifications and Standards'.

167 Lawrence Alexander, 'Scalar Properties, Binary Judgments', *Legal Studies Research Paper Series*, no. Research Paper No. 07-19 (October 2005).

limit) the contours of the reality that the system can apprehend. Under data-driven AI systems, the categories are not always delineated in advance, but can be suggested by the system itself based on patterns it identifies in the data it is fed. The contours of the system's apprehended reality hence hinge upon the patterns it may – or may not – pick up, and the way in which it categorizes data.[168]

Two caveats should be made regarding the digitalization of such data categorization, analysis and decision-making. Firstly, if we wish to process data in computerized form, we typically first need to translate the abovementioned concepts into abstract numbers that represent them, ultimately based on a binary system of zeroes and ones.[169] This translation from concept to code already entails a first risk of meaning getting lost in translation, and puts the person responsible for the translation in a position of power – albeit a hidden one. Secondly, the digitalization of the process – and its more mathematized outlook – may render it easier to forget the partiality and limitations of the concepts we use in the first place. We might hence be gaining speed through more efficient data computation, while risking to lose the existential plurality, relationality and holisticness of the elements behind they represent. Furthermore, the more data is gathered, the more confidence we tend to have in the computations, even if – as stressed previously – these solely rely on the choices of human developers.

In light of the societal paradigm discussed in Chapter 3, AI-enabled data analysis regarding the domain of human action is only possible if we treat such action – and the humans behind it – as numerical patterns that can be scientifically analyzed, all the while knowing that the diversity of the human condition does not lend itself to such reduction. Already before the advent of AI, Arendt criticized the "mathematical treatment of reality"[170] through statistics, which she linked to the victory of 'society' over 'the public'[171] – a development characterized by the conformity of individuals at the cost of their diversity, due to a lack of participation to the public realm in which it is possible to have an open, political dialogue.[172]

Simultaneously, individuals – or their features – that are not covered by the concepts which are codified into the system, or by the model that the algorithm identifies based on data patterns, are statistical outliers. When statistical rationality becomes prevalent in the intersubjective sphere, unfortunately, those outliers also risk being left 'out' in reality. Consider the example of IBM's facial recognition system, which had poorer accuracy scores for the identification of non-white-males. After much criticism, the company aimed to increase the 'diversity' of its facial

---

168 See in this regard also Crawford, *Atlas of AI*, 127.

169 Consider in this regard Laurence Diver, 'Interpreting the Rule(s) of Code: Performance, Performativity, and Production', *MIT Computational Law Report*, 15 July 2021, https://law.mit.edu/pub/interpretingtherulesofcode/release/1. See also Dufour, *Les mystères de la trinité*, 26.

170 Arendt, *The Human Condition*, 43.

171 See in this regard Arendt, *The Human Condition*, 28 and following.

172 She captures this as follows: "*To gauge of the extent of society's victory in the modern age, its early substitution of behavior for action and its eventual substitution of bureaucracy, the rule of nobody, for personal rulership, it may be well to recall that its initial science of economics, which substitutes patterns of behavior only in this rather limited field of human activity, was finally followed by the all-comprehensive pretension of the social sciences which, as 'behavioral sciences,' aim to reduce man as a whole, in all his activities, to the level of a conditioned and behaving animal*" in Arendt, *The Human Condition*, 45.

recognition data and to enhance the accuracy of its results in order to make the system 'fairer'.[173] As Crawford notes, "*though well intentioned, the classifications that they used reveal the politics of what diversity meant in this context. For example, to label the gender and age of a face, the team tasked crowdworkers to make subjective annotations, using the restrictive model of binary gender. Anyone who seemed to fall outside of this binary was removed of the dataset*."[174] However, as Arendt stresses, precisely these outliers – namely the individuals, events, traits – which do not necessarily fall under an identified pattern of conformity, are the very things that can make life meaningful.[175] The fact that they do not fit into the binary rationality of the system does not alter this, but risks obliterating their meaningfulness.[176]

As long as we rely on those systems outside the intersubjective realm – for instance, to optimize a production process in a factory – this need not pose grave concerns. Yet importing the same logic into the very core of our social domains requires applying mathematical rules to things that cannot truly be grasped thereby and leads to the reductionist paradigm described above.[177] In sum, the efficiencies we gain by 'rationalizing' human processes in a technocratic manner[178] have a risky counter-part, as they also entail a literal 'systematizing' and hence 'totalizing' of the process.

---

173 Crawford, *Atlas of AI*, 132.

174 Ibid.

175 The ability and meaningfulness of 'being' an outlier, and to deviate from the logic of determined lines, can also be linked to the discussion of the *declination* in Karl Marx' doctoral dissertation (Karl Marx, *Differenz Der Demokritischen Und Epikureischen Naturphilosophie - Doktordissertation (1841)* (Hofenberger, 2014). This discussion deals with the difference in perspectives between Democritus and Epicurus about atomic theory. Epicurus, in deviation from Democritus' view that atoms move in a straight line in space (which ultimately results in a deterministic world view) introduces the theory that atoms instead fall perpendicularly. This opens up the possibility for atoms to undergo a small declination in their movement, which can cause them to fall outside of the determined movement, and even to make new connections. Marx links this declination to the possibility for human freedom in a deterministic world. See in this regard also Luc Anckaert, 'The Thunderbolt of Evil and Goodness without Witnesses: In Conversation with Vasili Grossman, Life and Fate', *Religija Ir Kultūra* 18–19 (2016): 34. We can apply this interpretation to the problem at hand, since excluding or ignoring 'outliers' within an AI system can ultimately lead to the exclusion of human freedom to deviate from the norm, even if, as Arendt clarifies, these outliers are precisely where meaning can be found.

176 Arendt puts it as follows: "*The laws of statistics are valid only where large numbers or long periods are involved, and acts or events can statistically appear only as deviations or fluctuations. The justification of statistics is that deeds and events are rare occurrences in everyday life and in history. Yet the meaningfulness of everyday relationships is disclosed not in everyday life but in rare deeds, just as the significance of a historical period shows itself only in the few events that illuminate it. The application of the law of large numbers and long periods to politics or history signifies nothing less than the willful obliteration of their very subject matter, and it is a hopeless enterprise to search for meaning in politics or significance in history when everything that is not everyday behavior or automatic trends has been ruled out as immaterial*." See Arendt, *The Human Condition*, 42–43.

177 Consider, in this regard, the example that was raised previously concerning the difficulty to translate the notion of a person's 'creditworthiness' or 'trustworthiness' to a utility function. AI systems might be able to process tons of data, yet this does not mean they are able to adequately capture these concepts mathematically.

178 See also Andersen, 'The Technocratic Rationality of Governance - the Case of the Danish Employment Services'.

(b) From little goodness to Goodness

As already noted in Chapter 2, the deployment of AI often stems from a desire to improve processes and decisions so as to increase human wellbeing.[179] In this regard, we can also recall the rise of 'AI4Good' initiatives, seeking to deploy AI to, for instance, advance the UN Sustainable Development Goals.[180] It is, however, essential to examine what precisely AI developers understand as an 'improvement', and which assumptions underlay this understanding. The fact that these developers are often private companies – even if the systems can be used in public contexts – also means that the decisions of what constitutes an 'improvement' are typically not subjected to a democratic debate. Decisions that bear normative or even political values are hence not only mathematized but also privatized[181], while their consequences apply at scale.

This intention to do 'good', in what can only be described as a systematic and potentially totalizing way of codifying this 'good' through AI, is reminiscent of Emmanuel Levinas' discussion regarding the 'little goodness' versus capitalized 'Goodness'. In this discussion, Levinas draws on the novel '*Life and Fate*', written by Vasily Grossman.[182] *Life and Fate* details events during the Second World War and provides (comparative) perspectives about the totalizing regimes of Nazism and Stalinism. A few characters in the novel – especially Ikonnikov, described as a 'holy fool' – showcase "isolated acts of senseless kindness",[183] which stand in stark opposition to the great totalitarian visions of "the Good".[184] Levinas recounts the significance thereof as follows:[185]

> Grossman's eight hundred pages offer a complete spectacle of desolation and dehumanization…. Yet within that decomposition of human relations, within that sociology of misery, goodness persists. There is a long monologue where Ikonnikov – the character who expresses the ideas of the author – casts doubt upon all social sermonizing, that is, upon all reasonable organization with an ideology, with plans…. Every attempt to organize humanity fails. The only thing that remains undying is the

---

[179] Furthermore, as noted in Chapter 4, and as evidenced by the abovementioned example of IBM's facial recognition system, AI developers are increasingly turning to technical fixes to improve – and enhance the 'fairness' – of their models. On the difficulty to define and ensure 'fairness' in the context of AI, and the virtual impossibility to eliminate biases with technical fixes, see e.g., Reuben Binns, 'On the Apparent Conflict Between Individual and Group Fairness', 14 December 2019, http://arxiv.org/abs/1912.06883; Alex Hanna et al., 'Towards a Critical Race Methodology in Algorithmic Fairness', *ArXiv:1912.03593 [Cs]*, 7 December 2019; Crawford, *Atlas of AI*.

[180] Josh Cowls et al., 'Designing AI for Social Good: Seven Essential Factors', *SSRN Electronic Journal*, 2019; Nenad Tomašev et al., 'AI for Social Good: Unlocking the Opportunity for Positive Impact', *Nature Communications* 11, no. 1 (18 May 2020): 2468; Bettina Berendt, 'AI for the Common Good?! Pitfalls, Challenges, and Ethics Pen-Testing', *Paladyn, Journal of Behavioral Robotics* 10, no. 1 (1 January 2019): 44–65; Ricardo Vinuesa et al., 'The Role of Artificial Intelligence in Achieving the Sustainable Development Goals', *Nature Communications* 11, no. 1 (13 January 2020): 233.

[181] See in this regard also Linnet Taylor, 'Public Actors Without Public Values: Legitimacy, Domination and the Regulation of the Technology Sector', *Philosophy & Technology*, 20 January 2021.

[182] Vasily Grossman, *Life And Fate*, trans. Robert Chandler (London: Vintage Classic, 2017).

[183] Morgan, *Discovering Levinas*, 18.

[184] Luc Anckaert, 'Goodness without Witnesses: Vasily Grossman and Emmanuel Levinas', in *Levinas and Literature*, ed. Michael Fagenblat and Arthur Cools (De Gruyter, 2020), 226.

[185] See in this regard also Morgan, *The Cambridge Introduction to Emmanuel Levinas*, 23.

> goodness of everyday, ongoing life. Ikonnikov calls that 'little act of goodness'…. This 'little goodness' is the sole positive thing…. [I]t is a goodness outside of every system, every religion, every social organization.[186]

The systematized Goodness – which can be applied at scale, yet bears a risk of totalitarianism – is explicitly rejected and, instead, it is the small-scale, intersubjective, 'little goodness' that is cherished.[187] Any utopian thought eventually reifies individuals into abstractions, thereby eliminating the face-to-face relationships that constitute the core of our ethical existence. This totalizing tendency is present in any broad system, and large-scale AI networks are unlikely to escape this risk, no matter how good the underlying intentions of the systems' developers.

One can, furthermore, link the underlying rationale of this phenomenon to Arendt's distinction between work and action.[188] Whereas action is an inherently open and unpredictable activity that human beings undertake together – with political deliberation between a plurality of actors as primary example – work is instead focused on the fabrication of things in a specific way, aimed to counter unpredictability and open-endedness.[189] When both types of activities are confused, things go astray – and that is precisely what happens, according to Arendt, in totalitarian contexts.[190] She warns that this confusion still plagues us today. By substituting action for work, the public realm loses its openness of views and is instead infused by a specific idea of how things should be, just like an artist has a specific idea of how a statue should look like before carving it out of stone.[191]

A different variation of this warning can be found in Zygmunt Bauman's analysis of modernity, in which he uses the metaphor of a wild garden which is 'civilized' by eliminating weeds that do not fit into the gardener's aesthetic view.[192] For Bauman, this type of rationality is reflected in the horrific events of the Holocaust, during which Jews and other marginalized populations were considered as a weed to be eradicated.[193] Drawing inter alia on Arendt's work, Bauman warns that the procedural rationality and taxonomic categorization of species which ran through the logic of Nazi Germany, can still affect us today.[194] Rationality that is unconstrained by morality and public deliberation – and omits intersubjectivity – can lead towards a dangerous

---

186 Emmanuel Lévinas, *Is It Righteous to Be?: Interviews with Emmanuel Lévinas*, ed. Jill Robbins (Stanford University Press, 2001), 89.

187 See also the discussion of the Little Goodness in Anckaert, 'The Thunderbolt of Evil and Goodness without Witnesses'.

188 Anckaert, 'Goodness without Witnesses', 226.

189 Arendt, *The Human Condition*, 143.

190 Foreword by Margaret Canovan, xxiii in Arendt, *The Human Condition*.

191 See Arendt, *The Human Condition*, 227. Arendt recalls that this is precisely the approach of Plato as regards the political realm. He had a clear pre-existing idea of the Good, and of the roles that each and every person in society should play in order to conform to this idea, which lead to a totalitarian view of the state – in the name of the Good – rather than the democratic ideal of the polis. See also Arendt, *The Human Condition*, 142.

192 Zygmunt Bauman, *Modernity and Ambivalence* (Cambridge: Polity Press, 1991). See also the discussion of Bauman's approach to 'Modern Rationality and the Camps' in Anckaert, 'The Thunderbolt of Evil and Goodness without Witnesses', 24.

193 Bauman's statement that the ancient wisdom *Quos Deus vult perdere, prius dementat* should be rephrased, since it appears that "when God wanted to destroy someone, He did not make him mad. He made him rational", summarizes this problem. See Zygmunt Bauman, *Modernity and the Holocaust* (Cambridge: Polity press, 1989), 142.

194 Bauman, *Modernity and the Holocaust*.

path.[195] The use of AI necessarily demands a codification and hence systematization of the 'Good' that the system should attain and, as we have seen, often non-transparently so. This leaves a lot of leeway – and hence power, also unconsciously or unwillingly – for AI designers to shape the system and hence to shape our world, which comes with a considerable responsibility and risk.[196]

(c) From 'You' to 'It'

Human rationality is closely related to language, through which we structure, organize and express our thoughts.[197] The way we use language, and the way we use language about AI, shapes the way we think. For Franz Rosenzweig, if we guide our thinking through reason alone, we risk reducing all that is to one single ground, and end up in Hegel's ontological monism. Instead, Rosenzweig therefore suggests to temper the reductive impact of reason through the faculty of speech – which forms a cornerstone of his aforementioned 'new thinking'.[198] Speech-thinking can be opposed to Logical-thinking, whereby the former assigns primacy to the relationship between the speaker and the person that is spoken to, rather than the relationship between the spoken word and that which it signifies.[199] We enter into relationships with each other through language, and specifically through the spoken word. Meaning hence arises through interaction with an alterity.[200] How we speak not only determines how we narrate reality, but also which types of relationships we engage in.[201] Interestingly, the importance of speech is emphasized by Arendt too. According to her, it is only through the shared ability of speech that humans can create a 'common world', which is, in turn, a precondition for the political life and for human action.[202] For Arendt, speech is hence an inherently political act.[203]

In contrast, in the algorithmized world, opportunities to engage in intersubjective relationships through speech are drastically reduced when the interlocutor is an AI system. Furthermore, when an individual is adversely impacted by an AI system – for instance because of a miscategorization, or even because of the absence of a category that fits her case – the

---

195 Kieran Flanagan, 'Bauman's Travels: Metaphors of the Token and the Wilderness', in *Liquid Sociology: Metaphor in Zygmunt Bauman's Analysis of Modernity*, ed. Mark Davis (Routledge, 2016), 61.

196 To provide a simple example, consider, for instance, the decision of AI-enabled recommender systems regarding which posts or news articles social media users get to see, and which posts are considered instead as 'not appropriate' or 'less relevant'.

197 See Anckaert, 'Franz Rosenzweigs Stern der Erlösung. Een hermeneutische en retorische benadering'.

198 Benjamin Pollock, 'Franz Rosenzweig', in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Spring 2019 (Metaphysics Research Lab, Stanford University, 2019), https://plato.stanford.edu/archives/spr2019/entries/rosenzweig/.

199 Anckaert, 'Franz Rosenzweigs Stern der Erlösung. Een hermeneutische en retorische benadering'.

200 As Pollock summarizes: "*At the center of this speech-thinking is a philosophy of dialogue which traces the awakening of selfhood through an I-You relation into which the self is called by the Absolute other*" in Pollock, 'Franz Rosenzweig'. Note that this absolute Other or God, is also the one who – in Levinas' writings – calls individuals towards the face-to-face encounter with the Other, which shows the strong influence of Rosenzweig on Levinas' work.

201 Anckaert, 'Language, Ethics, and the Other between Athens and Jerusalem. A Comparative Study of Plato and Rosenzweig', 545.

202 Peter J Verovšek, 'Integration after Totalitarianism: Arendt and Habermas on the Postwar Imperatives of Memory', *Journal of International Political Theory* 16, no. 1 (1 February 2020): 6.

203 In Arendt's words: "Wherever the relevance of speech is at stake, matters become political by definition, for speech is what makes man a political being". See Arendt, *The Human Condition*, 3.

possibilities for re-interpretation, contestation and adaptation are close to zero. First, the individual already needs to know she is being subjected to a classifying system. Second, she needs to realize – and potentially prove – the system's error. Third, there is often no other human being to enter into a relationship with, and to explain to what went wrong. Instead, the interlocutor is an AI system and relies on *syntax* rather than *semantics.*[204] AI systems do not actually 'understand' the meaning behind the patterns they identify, the categories they propose or the decisions they recommend. This also means that, for the affected individual, there is no opportunity to 'speak' or 'reason' with the system. One can only try doing so with the human developer or deployer of the system, who is not always easily accessible. Furthermore, merely adding a 'human in the loop'[205] to the system will not be of much help if that human uncritically[206] refers back to the system[207] – resourcefully captured by the comical sentence "*computer says no*"[208].

When considering Arendt's account of modernity, and her description of the way in which speech is losing ground to statistics, one can observe that speech has, in fact, lost its power not only figuratively but also literally. Figuratively, the "*sciences today have been forced to adopt a 'language' of mathematical symbols which, though it was originally meant only as an abbreviation for spoken statements now contains statements that in no way can be translated*

---

[204] See e.g. Searle, J. R., 'Minds, Brains, and Programs', reprinted in John Haugeland (ed.), Mind Design: Philosophy, Psychology, Artificial Intelligence, MIT Press/Bradford Books, Cambridge, Massachusetts, pp. 282-306, 1980. This view is not shared by all. For an overview of critiques on this view – and in particular on the 'Chinese Room argument' in which Searle conceptualizes this limitation, see Cole, David. 'The Chinese Room Argument'. In The Stanford Encyclopedia of Philosophy, edited by Edward N. Zalta, Winter 2020. Metaphysics Research Lab, Stanford University, 2020. https://plato.stanford.edu/archives/win2020/entries/chinese-room/.

[205] High-Level Expert Group on AI, 'Ethics Guidelines for Trustworthy AI'.

[206] Of course, the uncritical application of a binary rule that carries adverse effects on individuals does not require AI systems. Also outside an AI-context (even intelligent) human beings can choose to revert to a written rule's authority to devoid themselves of the responsibility to critically reflect on the consequences of their actions – despite the rule engendering blatantly disproportionate effects. Consider in this regard the not-so-hypothetical rule of rejecting a Master paper, not for any aspect of content or quality, but because it is two pages too long, and consequently impede that Master student from graduating for that very reason. For a further discussion of the uncritical application of rules by human beings (and how the use of AI systems can exacerbate the issues arising therefrom), see also Hannah Arendt's reflections on Eichmann's 'thoughtlessness' and Stanley Milgram's assessment of people's obedience to authority, which are further dealt with under Chapter 5.2.

[207] The fact that people tend to suffer from automation bias further aggravates this problem. See in this regard, for instance, Kate Goddard, Abdul Roudsari, and Jeremy C. Wyatt, 'Automation Bias: Empirical Results Assessing Influencing Factors', *International Journal of Medical Informatics* 83, no. 5 (1 May 2014): 368–75; J. Elin Bahner, Anke-Dorothea Hüper, and Dietrich Manzey, 'Misuse of Automated Decision Aids: Complacency, Automation Bias and the Impact of Training Experience', *International Journal of Human-Computer Studies* 66, no. 9 (September 2008): 688–99; Daniel Varona, Yadira Lizama-Mue, and Juan Luis Suárez, 'Machine Learning's Limitations in Avoiding Automation of Bias', *AI & Society* 36, no. 1 (March 2021): 197–203.

[208] This catchphrase first appeared in a comical sketch of the TV series Little Britain, where receptionist Carol Beer responds to customers' enquiries or requests by typing them into her computer and answering "Computer says no". It is emblematic of both the limitations of digitalization and the unwillingness of human beings responsible for the digitalized processes to remedy the situation, with the computerized process as excuse. See also Ahmad Alwosheel, Sander van Cranenburgh, and Caspar G. Chorus, '"Computer Says No" Is Not Enough: Using Prototypical Examples to Diagnose Artificial Neural Networks for Discrete Choice Analysis', *Journal of Choice Modelling* 33 (1 December 2019); Agnieszka Werpachowska, '"Computer Says No": Was Your Mortgage Application Rejected Unfairly?', *Wilmott* 2020, no. 108 (2020): 54–61.

back into speech."[209] This is, for her, a reason to "*distrust the political judgment of scientists qua scientists*", since they "*move in a world where speech has lost its power*."[210] Yet in the algorithmized world, speech has also lost its power literally, since the individual subjected to an AI system has no recourse to her ability to speak and take action, given that she faces a machine rather than a fellow human being. The *I-You* relationship makes way for an *I-It* relationship instead. Furthermore, since the representation of people in terms of numerical abstractions to analyze and predict their features strips them from their corporeality, turning *I/You* into *It*,[211] we ultimately risk ending up in an *It-It* relationship, leading to a literal de-humanization.[212] While this literal dehumanization process is a necessary part of digitalization, given the abovementioned risks, it can also lead to a figurative dehumanization, which is significantly worse.[213]

Along with the risk of the *dehumanization* of human beings, we can also observe another development, namely the *humanization* of AI. Anthropomorphizing approaches to AI are increasingly widespread, not only by sensationalist media seeking readers or by AI developers seeking funding.[214] A growing number of academics are seriously discussing – and even arguing for – the need to assign moral and legal rights to AI systems.[215] Evidently, this way of thinking about AI – which is increased through anthropomorphizing language, has an impact on human thinking more generally, as ethics discourse is language-based.[216] If this occurs without due caution, it risks further eroding the responsibility of the human beings behind the system. Indeed, if the AI system is biased, causes harm or infringes people's privacy, the system can be blamed instead of its developers – despite the manifest fact that the system hinges entirely on human choices. The erosion of this human responsibility in light of AI's humanization can hence exacerbate the identified problems.

Evidently, the above does not imply that every deployment of AI is problematic *per se*. As previously stressed, there is no need to deny the benefits that AI-enabled data analysis can generate, in a variety of contexts. Yet the point that the aforementioned thinkers make regarding the technocratic rationality that drives the mathematization of intersubjective phenomena, also

---

209 Arendt, *The Human Condition*, 4.

210 Ibid.

211 Though writing before the events of the second World War, of relevance to this point is also the previously cited work of Buber, *I and Thou*.

212 Indeed, individuals lose their corporeality by being *datafied* and reduced to abstract numbers that can be categorized and essentialized through the conceptual distinctions assigned to them. They are no longer seen as particular individuals, but as entities that correlate with other entities, and that have a certain probability of correlating with yet another set of entities.

213 Scott H. Hawley, 'Challenges for an Ontology of Artificial Intelligence', *Perspectives on Science and Christian Faith* 71, no. 2 (2019): 83–95.

214 Salles, Evers, and Farisco, 'Anthropomorphism in AI'.

215 See for instance Joshua C. Gellers, *Rights for Robots: Artificial Intelligence, Animal and Environmental Law* (Routledge, 2020); David Gunkel, *Robot Rights* (MIT Press, 2018). Their arguments do not necessarily rely on legal efficiency (as is e.g. the case for arguments concerning the allocation of legal personality to companies, who in this manner can assume economic responsibility) but instead hinge, for instance, on the fact that human beings increasingly treat AI systems *as if* they are human beings, in light of their inherent propensity to anthropomorphize inanimate objects.

216 Anckaert, 'Language, Ethics, and the Other between Athens and Jerusalem. A Comparative Study of Plato and Rosenzweig', 545.

applies to the underlying paradigm that drives AI's adoption. Lest we repeat the excessively 'rational' approaches to society's organization from the past, we need to be aware of the fact that the deployment of AI entails a different way of thinking and of approaching reality than the way in which we deal – or ought to deal – with each other when we prioritize the meaningfulness of human relationships.

5.2 Alterity – Algorithms and Banality

The way a society deals with alterity – or with that which is other – is emblematic of its values. As we have seen above, one of the main points of critique that were uttered by Rosenzweig, Levinas and Arendt on Western philosophy, each from their own perspective, focused on its reduction of alterity to one overarching system, whether through a Hegelian notion of the 'Absolute', or through the Heideggerian *Dasein* that reduces the 'other' to 'Being' and hence remains within the singular, or – where philosophy meets politics – through concrete totalitarian regimes. The 'other' ruptures philosophical 'totality'.[217] Each of these thinkers – and particularly Levinas and Arendt, who lived through World War II – were also personally confronted with what it means to be 'other' in the world, which inevitably influenced their writings.

For Levinas, our social existence coupled with the uniqueness of each human being, leads to the fact that we always encounter, within our human experience, something that is irreducibly other. This other is not an infinite God or a form of the Good, but a particular person that you stand face to face with.[218] "*L'absolument Autre, c'est Autrui*."[219] The other interrupts the narcissistic ego, calls it into question, and appeals to it – thus demanding its responsibility.[220] Levinas has undoubtedly been inspired by Rosenzweig's *Speech-thinking* and emphasis on relationality, since for the latter, the irreducible entities *God*, the *world* and the *human* can be opened up precisely in relation to each other and in being for the other.[221]

Also Arendt is conclusive on the essential role of others for the human condition. According to her, the other is not another 'I', but a unique individual in his or her alterity. She connects the togetherness of unique human beings not only with the ability of speech, but also with the activity of action – which is closely linked to the political. [222] Action and speech require a

---

217 Susan Handelman, 'Facing the Other: Levinas, Perelman and Rosenzweig', *Religion & Literature* 22, no. 2/3 (1990): 63. Handelman cites the discussion of the term 'panim' by Maimonides, arguably the most prominent Jewish philosopher and scholar of the Middle-Ages in '*The Guide of the Perplexed*' (12th century). See Moses Maimonides, *The Guide for the Perplexed*, trans. M. Friedländer, 4th ed. (New York: E. P. Dutton & Company, 1904), 16 and 53.

218 Morgan, *The Cambridge Introduction to Emmanuel Levinas*, 3.

219 Levinas, *Totalité et Infini - Essai Sur l'exteriorité*.

220 The Hebrew word 'panim' or face, as already used in rabbinic tradition, has the root 'panah' which implies a 'turning' towards something or 'aim', as well as 'attention or regard', which also clarifies Levinas' use of the term as a facing 'relation'. See Handelman, 'Facing the Other', 63.

221 Handelman, 64 and 72. See also Cohen, R. A., 'The Face of Truth in Rosenzweig, Levinas and Jewish Mysticism', pp. 175-201, in D. Guerrière, Phenomenology of the Truth Proper to Religion, Albany, SUNY Press, 1990.

222 In Arendt's words: "*All human activities are conditioned by the fact that men live together, but it is only action that cannot even be imagined outside the society of men. Action alone is the exclusive prerogative of man;*

common space in which a plurality of human beings can meet, exchange views, deliberate, and seek agreement on how to deal with political questions. This 'common world' or 'public realm', as she calls it, is the founding dimension of human experience which she calls 'worldliness': the sharing of a world in which individuals "*can meaningfully assume and express individuality, act, and interpret their political experiences*".[223] Most importantly though, while alterity in the form of other human beings is always there, the common space that humans need to interact with others and to have meaningful experiences is not something to be taken for granted.[224] It needs to be built together and it can be undermined – as was the aim, according to Arendt, of the totalitarian regimes she describes.[225]

When considering how intersubjective alterity fares in the algorithmized world, at least three issues can be noted, namely the fact that AI systems can be used to polarize (a) and isolate individuals (b), and that their deployment can banalize the ethically problematic decisions they enable (c).

(a) Polarization

It is through our being-in-the-world-together that significance can occur. Conversations with the Other – whether through texts, speech or language more generally – open up the possibility for new meanings to arise. This centralizes both the presence of the Other and the need for a dialogue with that Other, requiring a common space to do so. We have seen above how Arendt conceptualizes this space as the "public realm" in which we can interact with each other, drawing inspiration from the Greek polis state.[226] Furthermore, she notes that "*everything that appears in public can be seen and heard by everybody and has the widest possible publicity*".[227] By engaging in public deliberation and looking at the same thing through a diversity of perspectives, we can "*see sameness in utter diversity*" and let "wo*rldly reality truly and reliably appear*".[228] In other words: it is by coming together in a plurality of views that we not only engage with each other politically, but that we can also ensure that our world-view actually corresponds with reality.

---

neither a beast nor a god is capable of it, and only action is entirely dependent upon the presence of others." See Arendt, *The Human Condition*, 22.

223 Matthew Sharpe, 'When the Logics of the World Collapse - Zizek with and against Arendt on "Totalitarianism"', *Subjectivity* 3, no. 1 (April 2010): 54.

224 "*Wherever people gather together, it is potentially there, but only potentially, not necessarily and not forever*" in Hannah Arendt, *The Human Condition* (University of Chicago Press, 2019), 199; See also Svetlana Boym, 'From Love to Worldliness: Hannah Arendt and Martin Heidegger', *The Yearbook of Comparative Literature* 55, no. 1 (2009): 106–28.

225 Hannah Arendt, *The Origins of Totalitarianism* (Penguin Classics (2017), 1951); Sharpe, 'When the Logics of the World Collapse - Zizek with and against Arendt on "Totalitarianism"', 54.

226 She describes this as the "*the organization of the people as it arises out of acting and speaking together*" rather than as a city-state in its physical location. Focusing thus on the idea behind the polis rather than the historical city state, she emphasizes how this shared space "*rises directly out of acting together*", and enables "*the reality that comes from being seen, being heard and, generally, appearing before an audience of fellow men*". This, according to her, also leads to human excellence – as it provides a public place and hence incentive for it. See Arendt, *The Human Condition*, 198. See also Sue Spaid, 'Surfing the Public Square: On Worldlessness, Social Media, and the Dissolution of the Polis', *Open Philosophy* 2, no. 1 (31 December 2019): 670.

227 Arendt, *The Human Condition*, 27.

228 Ibid.

Arendt contrasts the public realm with the private and the social realm, and finds these spaces to have become increasingly – and problematically – blurred. She considers that the public realm is where human beings can enjoy their freedom and engage with each other on an equal footing to discuss matters of public interest.[229] In contrast, the private realm is dictated by private interests and by one's role and position within the household, requiring a uniform approach rather than a pluralistic one. Finally, the social realm is "*neither private nor public*" but came into existence by importing the uniform house-keeping idea of the private realm into the public realm. In the social realm, "*the scientific thought that corresponds to this development is no longer political science but 'national economy' or 'social economy'*".[230] Yet this development occurred at the cost of the political, since rather than harnessing a plurality of views, society turned into one super-household, with one interest. Where the social takes over the public realm, a mass society arises[231], characterized by two tendencies. The first concerns *loneliness*, as the loss of a shared public realm also entails the loss of meaningful others. The second concerns *conformity*. Individuals can no longer affirm their freedom by engaging with each other in a public sphere and appearing before an audience of fellow human beings. Instead, they conform to groups, mirroring the model of the unified household where only one opinion can be tolerated.[232] In the social realm, the free action of the polis is thus replaced by 'expected behavior' imposed by society and societal groups, excluding spontaneous action. To counter the oppression of societal conformity, Arendt appeals to restore the public realm as an essential shared space for human action.[233]

The problem of forced conformity can also be linked to the triangular relationship that was already evoked above. I cautioned that, in the dialogue between *I* and *You*, this *You* ought not to be reduced to an *I* (as much of Western philosophy was criticized for doing), nor should *You* be reduced to an *It* (which risks occurring when human beings are instrumentalized and essentialized or – in the context of AI – numericized and datified). Yet Arendt also warns against trying to reduce *You* to a *We*. The aim of an open dialogue with the other is not to unify the diversity of views into a multiplicity of the same. Respecting the other's alterity means engaging with a plurality of views, and maintaining the *I-You* relationship without reductions. With this in the back of our mind, let us now look at the shape of the public realm today.

The primary observation we must make, concerns the rise of the Internet and in particular online social media platforms, which for a large part now constitute the 'public realm' where we engage with each other. On these platforms, we talk with our friends and family; connect with groups of people who share our interests; read the news and update ourselves about national and global events; shop; engage with political content; and, more generally, co-shape public opinion. The

229 Arendt, *The Human Condition*, 30.
230 Arendt, *The Human Condition*, 28.
231 Spaid, 'Surfing the Public Square'.
232 See also Spaid, Surfing the Public Square', 669.
233 Arendt has, unsurprisingly, been criticized for romanticizing the ancient polis without elaborating on the problems that this historical phenomenon entailed. Yet if we take her account not as a historical description but as a philosophical discourse about an ideal, we can rely on the important insights this provides us, which are of great relevance still today. See also Sharpe, 'When the Logics of the World Collapse - Zizek with and against Arendt on "Totalitarianism"'.

global COVID-19 pandemic has further spurred our reliance and dependence on this online world. The advent of social media seemingly enlarged the opportunity for mutual engagement with each other, yet at the same time, given its underlying business model, it also foreclosed it. In theory, the common space in which one can be heard by fellow human beings has never been larger and more accessible. At the start of this phenomenon, many scholars therefore eagerly mused the possibilities for democratic deliberation and participation that would be opened up through these online platforms.[234] Yet the way in which this online common space is shaped has essentially undermined that hope and, to a large extent, the AI systems that are part of the platforms' infrastructure have contributed to such undermining.

Given the overly large amount of information posted on social media, AI systems are widely deployed to organize and prioritize the content we get to see. Based on the values for which the system is optimized, there are messages that appear at the top of our screen or rather at the bottom – or messages that we do not get to see at all.[235] Since extreme and polarizing content typically attracts more engagement than nuanced content, it is the former that often gets prioritized by AI systems – to the detriment of social solidarity. Furthermore, rather than confronting social media users with a plurality of views, individuals are instead often confronted with content they already like and are certain to engage with, thereby confirming their former – even if potentially one-sided or plainly wrong – views, which ultimately risks creating echo chambers rather than an open dialogue.[236]

Importantly, these echo chambers also allow public figures – such as politicians – to tailor their messages to their audience, and to no longer face public accountability for potentially problematic utterances, or for delivering entirely different messages to different people. 'Social' media would, based on Arendt's perspective, hence deserve its name. Whereas a truly 'public realm' puts a spotlight on everything that happens in public, AI-driven social media platforms can help those in power select what you get to see. Accordingly, public discourse can become more fragmented and shaped towards the entrenchment of existing power relationships, which

234 See Spaid, 'Surfing the Public Square'.

235 AI systems are typically designed to prioritize content that social media users will engage with, since more engagement means more revenues from advertisers, who are sustaining the social media business model given that access to social media is 'free'. Indeed, individuals typically do not pay a monetary sum to have access to online platforms, and can hence not truly be seen as their customers. Instead, they are more akin to a product, since platforms greedily use their personal data to better tailor content and advertisements to their preferences, and more specifically, to what they will engage with – regardless of whether the content of the message is newsworthy or truthful.

236 See e.g. Martin et al., 'From Echo Chambers to "Idea Chambers"'; Brent Kitchens, Steven L. Johnson, and Peter Gray, 'Understanding Echo Chambers and Filter Bubbles: The Impact of Social Media on Diversification and Partisan Shifts in News Consumption', *MIS Quarterly* 44, no. 4 (December 2020): 1619–49; Andrei Boutyline and Robb Willer, 'The Social Structure of Political Echo Chambers: Variation in Ideological Homophily in Online Networks', *Political Psychology* 38, no. 3 (2017): 551–69. Some studies have, however, nuanced these findings and suggested that users also get to see material from the other side of the political spectrum. See e.g. Seth Flaxman, Sharad Goel, and Justin M. Rao, 'Filter Bubbles, Echo Chambers, and Online News Consumption', *Public Opinion Quarterly* 80, no. S1 (1 January 2016): 298–320.

can contribute to the polarization of society and undermine intersubjective solidarity. If a *You* cannot be turned into a *We*, there may be no space for her.[237]

(b) Isolation

At the same time, AI systems also enable to surveil and collect ever more data from individuals, at ever-larger scale, to map their psychographic profiles with ever more details. These profiles can not only be sold to other parties who may have an interest in such information – from insurance companies to political parties – but can also be used to subliminally manipulate individuals based on psychological traits they may not even be aware of.[238] In addition, the AI-enabled collection and analysis of information can be used to monitor and stifle individuals who do not share the view of those in power, thereby preventing individuals from challenging certain views and from organizing themselves to better protect public interests – leading to their isolation.[239] Importantly, this risk is not only present in the online sphere. AI-systems are also becoming ubiquitous in physical public spaces. Consider, for instance, the introduction of AI-enabled facial and object recognition cameras, which likewise facilitate the automated and widespread tracking of individuals, and can thereby lead to chilling effects.[240]

The dangers of isolation, which runs precisely counter to a shared world, were highlighted by Arendt not just in *The Origins of Totalitarianism*, but also in *The Human Condition*, precisely in the chapter where she describes the importance of – having a space for – (political) action:

> Montesquieu realized that the outstanding characteristic of tyranny was that it rested on isolation – on the isolation of the tyrant from his subjects and the isolation of the subjects from each other through mutual fear and suspicion – and hence that tyranny was not one

---

237 Consider in this regard the example of Twitter's biased photo cropping algorithm. Pictures posted by Twitter-users are often too big to be shown in their entirety on the Twitter-feed, and are hence automatically cropped. To decide which part of the picture constitutes the picture's focal point and should hence be kept, and which part should be cropped away, Twitter used an algorithm to decide on the saliency of the picture's different aspects. Studies revealed that, when a picture shows different individuals, the cropping algorithm assigns a higher 'saliency' score to people with lighter skin tones, a slimmer appearance and younger age. This exemplifies in a very blunt manner how people that deviate from what the algorithm – based on the data it was trained on – considers as salient and hence optimizes for, can literally be cropped away. After having received complaints from users in late 2020, Twitter investigated the matter and published a study confirming the fact that its algorithm was biased in May 2019. It decided to no longer use it. See Rumman Chowdhury, 'Sharing Learnings about Our Image Cropping Algorithm', Twitter, 19 May 2021, https://blog.twitter.com/engineering/en_us/topics/insights/2021/sharing-learnings-about-our-image-cropping-algorithm. Moreover, recently, Twitter organized an 'Algorithmic Bias Challenge', which invited hackers to identify other problems in Twitter's cropping algorithm. The winner of the challenge brough to light that, in addition to racial bias, the algorithm also seemed to have an age and weight bias. The study, openly reported on Github, can be found here: https://github.com/bogdan-kulynych/saliency_bias.

238 Liesl Yearsley, 'We Need to Talk about the Power of AI to Manipulate Us', MIT Technology Review, accessed 12 November 2019, https://www.technologyreview.com/s/608036/we-need-to-talk-about-the-power-of-ai-to-manipulate-humans/. Consider also Zuboff, *The Age of Surveillance Capitalism*.

239 Council of Europe Ad Hoc Committee on Artificial Intelligence (CAHAI), 'Feasibility Study'.

240 In China, such systems are used in public streets to ensure people's conformity with the governments' rules and views, as well as in other public and private spaces – from schools to shopping malls. Yet also in Europe, the use of these systems in public spaces is increasing. This further drives the imposition of conformity – directly or indirectly – over freedom and plurality, and risks turning everyone who does not conform into an outcast. The threat of being surveilled risks preventing people from gathering, speaking and taking action, leaving them dispersed and undermining solidarity. See AlgorithmWatch, 'Automating Society Report 2020'.

> form of government amongst others but contradicted the essential human condition of plurality, the acting and speaking together, which is the condition of all forms of political organization.[241]

The passage stresses the essentiality of plurality, as well as drawing attention to the role of power – and more precisely political power – that can be strengthened or diminished by virtue of this essentiality. Tyranny is here linked to a space where human beings lost their capacity to act and speak together in a way that does justice to their plurality. It is here that, once again, reference can be made to the risks related to the algorithmic processes of our online space, which can be used in a way that undermines a plural public discourse and instead can isolate and atomize individuals. This atomization is a precondition for the success of totalitarian regimes, since it tends to drive people towards more totalitarian movements.[242] By creating mistrust instead of trust, and by blurring the line between fact and fiction, the shared reality of the public realm is undermined, and with it, the empowerment of individuals to challenge power.[243]

It is essential to underline that these trends are not caused by AI systems *per se*. As abundantly stressed above, AI systems are designed, developed and deployed by human beings. Hence, the consequences of the use of these systems can be traced to the humans behind them – who can also choose to use these systems in different non-damaging ways. However, the technology expediates these practices and thereby risks exacerbating problematic behavior. While human beings can equally collect information from individuals manually, or with basic technological tools (as they have been doing in the past), AI systems enable them to do so at a much larger scale, and hence increase the possibility for mass-manipulation and mass-surveillance.[244] In sum, without due care, the algorithmized world can facilitate the very tendencies that run counter to the idea of a 'common world', and might ultimately even normalize them.

(c) Banalization

In addition to the risks of polarization and isolation, which put intersubjective relationality under strain, there is another problem that needs to be mentioned. I already noted previously the ubiquitous substitution of human-human interaction with human-machine interaction, which risks eliminating the possibility to engage in a dialogue with fellow human beings and co-create meaning. In a very concrete way, the face of the other as conceptualized by Levinas disappears, and in its place appears a computer screen. Unlike the face of the other, this computer screen does not appeal to us or instill us with an inherent and inescapable responsibility. Accordingly, this loss can affect an essential element of our – fundamentally ethical – human condition.[245] It

---

241 Arendt, *The Human Condition*, 202.

242 Matthew Sharpe, 'When the Logics of the World Collapse - Zizek with and against Arendt on "Totalitarianism"', *Subjectivity* 3, no. 1 (April 2010): 53–75.

243 Boutyline and Willer, 'The Social Structure of Political Echo Chambers'; Spaid, 'Surfing the Public Square'.

244 Yeung, 'Responsibility and AI - A Study of the Implications of Advanced Digital Technologies (Including AI Systems) for the Concept of Responsibility within a Human Rights Framework'.

245 There have been attempts by some scholars to apply Levinas' teachings regarding the appeal of the face of the other to 'AI systems'. According to these scholars, such systems – especially, but not exclusively, when built in an anthropomorphic way – might create such an appeal towards us too, despite their non-human nature, and force us to take up responsibility for their being, for instance by granting them moral and/or legal rights. For reasons of space, I will not engage here with such approach. I can refer the interested reader to, e.g., Benjamin

is the design of the system that will henceforth delineate the contours of meaning that can arise from the human-computer interaction, pushing it into a straitjacket that suits the AI designer, or the organization that pays for the system's design.

Let us now focus on the impact that this altered relationship can have not on the individual subjected to the AI system, but on the individual who deploys it, by venturing into one of Arendt's not yet above-cited works, namely her *Lectures on Kant's Political Philosophy*.[246] In it, she develops more philosophically the concept of 'the banality of evil' which she introduced in her report on Adolf Eichmann's trial.[247] She ascribes the deeds of Eichmann, in charge of the logistics of the mass deportations of Jews during World War II, not as monstrous but as *banal*, and arising out of 'thoughtlessness' rather than out of a radically evil inclination – an idea that was coined as the banality of evil. During his trial, Eichmann stated he was simply doing his job and following the orders of his superiors, and even went as far as quoting Kant's categorical imperative to explain his deeds. Arendt understands this as an ethical thoughtlessness arising from a rationally constructed world, where an imposed order is prioritized over any moral responsibility. Eichmann, to put it simply, did not *think*. The reason for this thoughtlessness can be found, according to Arendt, in the lack of his ability to test his actions against an intersubjective or common judgment. Drawing on Kant's *Kritik der Urteilskraft*, she specifies that such judgment should meet three criteria: a comparison with the judgments that others might have, a way of thinking that displaces one's thoughts to the other's thoughts, and conformity or consistency with itself.[248] The fact that Eichmann failed to take a common perspective to judge his actions, based on consideration for and with others, led to his immoral deeds, and worse, to his inability to perceive them as such.

This philosophical account of *thoughtlessness* was to a large extent empirically corroborated by Stanley Milgram, who explicitly referred to Arendt's work and carried out an experiment to assess how obediently people acted under authority.[249] Individuals were asked to administer increasingly high electric shocks to a volunteer, whenever that volunteer answered erroneously. While the shocks were fake, the grim results of his experiment indicated the high rate of individuals who, sitting behind a machine and faced with the choice to obey to authority or refrain from hurting another human being (even upon that human being's specific request to stop), all too often opted for the former. Milgram analyzed the results of the – multiple variations of his – experiment and drew a number of conclusions that are relevant for our purpose. Thus he notes that "*distance, time and physical barriers neutralize the moral sense*."[250] The further

S. Wohl, 'Revealing the "Face" of the Robot: Introducing the Ethics of Levinas to the Field of Robo-Ethics', in *Mobile Service Robotics* (17th International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines, Poznan, Poland: World Scientific, 2014), 704–14; David J. Gunkel, 'The Other Question: Can and Should Robots Have Rights?', *Ethics and Information Technology* 20, no. 2 (1 June 2018): 87–99.

246 Hannah Arendt, *Lectures on Kant's Political Philosophy* (Chicago: University of Chicago Press, 1982).

247 Hannah Arendt, *Eichmann in Jerusalem: A Report on the Banality of Evil* (Viking Press, 1963).

248 Anckaert, 'The Thunderbolt of Evil and Goodness without Witnesses', 26.

249 Stanley Milgram, *Obedience to Authority: An Experimental View* (New York: Harper Perennial, 2009).

250 Milgram, 157.

away the individual was from the volunteer subjected to the shock, the higher the obedience rate. [251]

Moreover, Milgram explains that individuals automatically adopt a number of internal mechanisms to cope with the tension they face in this ethically difficult situation. One of those mechanisms concerns the divestment of moral responsibility by pointing towards a hierarchical higher authority – much like Eichmann tended to do. Another consisted of developing the opposite tendency of anthropomorphism, by attributing impersonal qualities to the human being that is hurt, making it easier to cope with their role as hurter. [252] Furthermore, Milgram describes "*the tendency of the individual to become so absorbed in the narrow technical aspects of the task that he loses sight of its broader consequences*".[253] The fact that the act becomes fragmented – the individual is no longer the person who decides to carry out the problematic act and is confronted with the direct consequences, but there is a chain of actions in between – likewise facilitates its execution.[254] Milgram therefore cautioned not only for the risk of malevolent authority, but in particular for the dehumanizing effect of these mechanisms or 'buffers' that divest individuals from their sense of responsibility.

When we now turn to our algorithmized world, we are faced with yet another set of concerns. First, AI systems can be said to go a step further than more traditional technologies precisely in their ability to also 'reason' and 'learn' based on the data they are provided with, and to take action on that basis, without the necessity for human intervention. The delegation of authority over certain decisions to AI systems is thus in principle deliberate. Second, similar to the process described by Milgram, there is often a physical distance between the person responsible for the AI system (the designer or deployer) and the individual subjected thereto. Indeed, the AI system is meant to take over tasks from human beings, providing them with the possibility to monitor these tasks from a distance rather than carrying them out themselves.[255] This physical distance facilitates an emotional distance from the individual subjected to the AI system, and from a feeling of responsibility in case that individual is harmed.

Third, the mathematical rationalization that the system engenders can also enhance this emotional distance. Individuals risk excessively relying on AI systems since their perceived objectivity, scientific nature and alleged high accuracy gives them an air of authority. This can

---

[251] Milgram explicitly refers to the set-up of the experiment, and the role that technology played therein: "*While technology has augmented man's will by allowing him the means for the remote destruction of others, evolution has not had a chance to build exhibitors against these remote forms of aggression to parallel those powerful inhibitors that are so plentiful and abundant in face-to-face confrontations.*" See Milgram, 157.

[252] Milgram, 8.

[253] Milgram, 7.

[254] Milgram calls this a dangerously typical situation in complex societies: "*it is psychologically easy to ignore responsibility when one is only an intermediate link in a chain of evil action but is far from the final consequences of the action. Even Eichmann was sickened when he toured the concentration camps, but to participate in mass murder he had only to sit at a desk and shuffle papers*." See Milgram, 11.

[255] We need not resort to the most evident yet perhaps also most radical example of autonomous weapon systems to illustrate this issue. Consider this problem in the context of facial recognition technology. Instead of planting a police officer on every street corner, an AI-enabled facial recognition system can monitor the street and search for an individual – or group of individuals – while the police officer, at the office, can be alerted by the system if someone is found. Given the system's remote operation, the individual might not even know that she is being filmed and identified.

be linked to the risk of automation bias, or the human "*tendency to disregard or not search for contradictory information in light of a computer-generated solution that is accepted as correct and can be exacerbated in time-critical domains*."[256] It is therefore with relative ease that individuals rely on the authority of AI systems, in particular given the systems' superior computational skills. This occurs even more in contexts of time pressure, when people do not have the time to double-check the system's suggestion, or in contexts of scarcity of information, when people lack the data or knowledge to assess the system's reliability. In such case, the hope we can vest in the intersubjective relationship – recalling Levinas' reference to *Life and Fate*, stressing that it is in relation of one human being to the other that goodness persists,[257] rather than in the context of a *systemic* approach to Goodness – gets even slimmer. Fourth, AI systems are often composed of different components that interact with each other within a broader network or chain, which further alienates the AI deployer from the consequences of the deployment and facilitates the evasion of responsibility. This gives rise to the difficulty of the *many hands*-problem,[258] which, in the context of AI, is only intensified by the opacity surrounding the different types of (interacting) conducts and systems.[259]

Finally, if we revisit Arendt's account of thoughtlessness due to a lack of adequate judgment, we can note the absence of a plurality of views and common deliberation regarding the parameters and optimization function that AI systems should have in the first place, as well as regarding the domains and conditions in which they should be used. This further problematizes the responsibility that should be taken by the AI developer or deployer – in addition to the more general undermining of a 'common world' by using AI systems in a potentially isolating and polarizing manner.[260]

We are hence faced with a situation in which the persons deploying potentially harmful AI systems may not be aware of the adverse impact caused by the system, or may deploy the system

---

256 Mary Cummings, 'Automation Bias in Intelligent Time Critical Decision Support Systems', *American Institute of Aeronautics and Astronautics*, 1 November 2014, https://web.archive.org/web/20141101113133/http://web.mit.edu/aeroastro/labs/halab/papers/CummingsAIAAbias.pdf.

257 Morgan, *The Cambridge Introduction to Emmanuel Levinas*, 23.

258 Dennis F. Thompson, 'Designing Responsibility: The Problem of Many Hands in Complex Organizations', in *Designing in Ethics*, ed. Jeroen van den Hoven, Seumas Miller, and Thomas Pogge, 1st ed. (Cambridge University Press, 2017), 32–56.

259 As noted elsewhere, in the context of the wide-spread use of AI systems, not one but three levels of this problem come to mind. First, at the level of the AI system, multiple components developed and operated by multiple actors can interact with each other and cause harm, without it being clear which component or interaction is the direct contributor thereof. Second, at the level of the organization or institution deploying the system (whether in the public or private sector), different individuals may contribute to a process in many different ways, whereby the resulting practice can cause harm. Last, this issue manifests itself at the level of the network of organizations and institutions that deploy the problematic AI application. The scale of these networks and their potential interconnectivity and interplay renders the identification of the problematic cause virtually impossible – even more so if there isn't necessarily one problematic cause. See Smuha, 'Beyond the Individual: Governing AI's Societal Harm'.

260 It is also futile to count on the AI system as such to exercise proper judgment, since it is banal by definition. AI systems cannot place themselves in the perspective of another person to verify whether their actions are justifiable from a 'common sense' judgment perspective, since they lack any common sense, and have no 'understanding' of the significance of their actions on other persons. They are, in essence, ideal bureaucratic tools, which will follow orders according to the way in which they are programmed.

with a problematic thoughtlessness that deprives them of any sense of moral responsibility for the harm done to others, whose face they do not see – literally nor figuratively. At the same time, the affected others, assuming they are aware of – and can prove – the fact that an AI system adversely impacts them, can find themselves in a situation in which they have no human being to turn to in order to challenge this impact, or to seek a shared space in which to have an open dialogue on how the system can respect their individuality and otherness. This comes in addition to the abovementioned dehumanization process associated with the 'translation' and classification of human actions and traits into abstract numbers and categories. In sum, if these concerns are not duly considered, our engagement with alterity in the algorithmized world may be difficult to reconcile with an intersubjective ethics.

5.3 History – Algorithms and Infinity

I now discussed how we think and how we treat others as part of the human condition, taking the perspective of intersubjectivity, and measuring this up against the reality of the algorithmized world. Strongly related to our conception of rationality and alterity is our experience of time and history, to which this last section is devoted. In his *Stern der Erlösung*, Rosenzweig pays considerable attention to the temporal situatedness of human beings[261] – a theme that also runs through Levinas' writings. Human experience necessarily has a temporal character, which spans over a past, present and future. From the perspective of the unique individual, time is however not experienced as the measurable, mathematical, scientific time, but rather as 'duration', as famously conceptualized by Henri Bergson, another Jewish philosopher.[262] Both Rosenzweig and Levinas followed Bergson in this distinction and give an account of human temporality without reliance on a scientific time. Yet whereas Rosenzweig connects this temporal experience to revelation – which for him is the ultimate human orientation point – Levinas instead seeks to ground this experience of time in the face-to-face relationship with the other person.[263]

Our temporal existence, according to Levinas, is enabled and given meaning through our encounter with the face of the other (which for him is the location of revelation) which summons us in its needfulness.[264] Time has thus an inherently intersubjective dimension, as it is "*impossible to speak of time in a subject alone, or to speak of a purely personal duration*".[265] This also means that our experience of time has an inherently ethical dimension. A brief note is needed here of Levinas' Messianistic eschatology, which is closely entwined with his approach to time and history. For him, eschatology does not concern the end of history, but the 'beyond' of history, which "*draws beings out of the jurisdiction of history and the future; it arouses them*

---

261 Pollock, 'Franz Rosenzweig'.

262 See Henri Bergson, Time and Free Will (1888).

263 As Michael Morgan explains: "*For Rosenzweig, revelation is the divine command to redeem the world through love; for Levinas, a similar sense of obligation to accept and help others and to alleviate their suffering is the content of the face to face. Just as revelation, then, gives time and history an absolute structure and direction – and a determinate future and goal, so does the face to face*". Morgan, *The Cambridge Introduction to Emmanuel Levinas*, 167.

264 Morgan, 167.

265 Levinas, *Le Temps et l'Autre*.

in and calls them forth to their full responsibility". Hence, by taking eschatology away from the conception of the end of time, it becomes centered on living each moment in function of the responsibility we have to care for each other and to alleviate the other's suffering. Life's meaning is constituted not of that which will eventually occur, but of how we live in the here and now, during each instance of time.[266] The future hence concerns that which we should do in the present, namely, acting in line with our ethical obligation towards the other. Furthermore, given its unknown and ungraspable nature, also the future is inherently other.

In his later works, Levinas focuses more on the past, and specifies that human existence occurs in relation to an immemorial past of ethical responsibility to the other person, and in anticipation of a realization of that responsibility. Living in a world with others, and living temporally, means what it does to us because of our obligations to serve the needs of those others.[267] Temporality – and particularly the past – is a subject that likewise occupied Arendt, who strongly emphasized the need for remembrance as a vital component of the maintenance of a 'shared world'. The relationship with the past should not be one of stories and myths like in nationalistic settings, but one that can be contested, and in which we must find elements to make politics for the present meaningful.[268]

What now should we make of this – alterity-oriented, ethics-infused and meaning-laden – approach to history in an algorithmized world? In the below, I discuss these questions in relation to the past (a), present (b) and future (c).

(a) Past

When we consider the way in which we experience the past in an algorithmized world, we are forced to make a striking observation: there is no real past. The time that lays behind us, and that we are trying to make sense of in the present, is never truly behind us. As noted above, AI systems rely on data to reason, learn, analyze, draw inferences and – on that basis – make suggestions, take decisions, or execute tasks. Yet the data they rely on concerns, by necessity, data from the past. This data from the past is hence modelled, processed and rehashed to provide outcomes for today and tomorrow. Each time new data is produced and added to the system, this is joined with the previous data from the past in order to adjust the algorithmic model. This process is repeated through a continuous loop of model optimization,[269] and has important consequences for our relationship with history.

First, the fact that past data is continuously used to prepare and maintain models for the present and future, means that it is almost impossible to break free from the problematic aspects that make up our history, and particularly from structural historical inequalities and systemic discriminations. Those with a position of power in the past can rely on AI to see that position strengthened, while those who were already in a vulnerable or marginalized position risk

---

266 It can be noted that this eschatological view of Levinas differs from the view of Rosenzweig, who instead conceives history as 'salvation-history', in which it is the sum of all human experience together that has meaning. See in this regard also Morgan, *The Cambridge Introduction to Emmanuel Levinas*, 170.

267 Morgan, 180.

268 Verovšek, 'Integration after Totalitarianism', 9.

269 David Theo Goldberg, 'Coding Time', *Critical Times* 2, no. 3 (1 December 2019): 353–69.

remaining entrenched therein.[270] The abovementioned example of Amazon's hiring algorithm that was biased against women is emblematic of this problem: the use of past (male) data to create a model of the ideal future candidate, will entrench the position of the male tech worker, to the detriment of women.[271] The same problem recurs as regards other inequalities, based on ethnicity, social class or disability. Given the scale at which AI systems can be used, historical discriminations risk not only being entrenched, but also expanded and exacerbated. The aforementioned opacity problems only worsen this risk.

Second, this continuous importation and perpetuation of the past also makes it difficult to detect and repair past wrongs. As long as the algorithmic paradigm is maintained, the challengeability of the AI system's processes and outcomes is arduous, as it goes against the optimistic spirit of progress that AI is promising us. Yet this optimism hinders the identification and acknowledgment of wrongs from the past, especially as those wrongs are often still reflected in society through the existing systems of power.[272] Since AI systems are not only technical artefacts, but part of a broader system of networks and institutions,[273] the individual subjected to the system risks to remain infinitely situated in the problematic and totalizing structures of the past, without an escape – unless such escape is explicitly created.

Importantly, the fact that the past remains part of the present is not problematic *per se*. Remembrance plays a vital role in the constitution of the human condition and in our intersubjective humanity.[274] Instead, the problem manifests itself through the lack of a shared conversation and vision – arising from that intersubjective humanity – regarding which elements of the past should be perpetuated, and under which conditions. In the algorithmic paradigm, this vision is determined by those developing and deploying the algorithm, rather than in a common world.

Third, the past also keeps chasing us into the present in a very individual way. We are, today, leaving ever more digital traces of ourselves. Unlike paper traces, those digital traces can be copied and stored rapidly and virtually infinitely, and fed into AI systems to prepare detailed personal profiles. Such profile can contain the comments you made on your *myspace* webpage when you were twelve, the grades you obtained at school, the pictures you posted on Facebook, the track you ran during your jogging excursion, the number of times you drove to the hospital and the advertisements you clicked on. These traces of one's past – digitized, analyzed and, potentially, sold and publicized – can always be brought back into the present (for instance to embarrass or blackmail you), as well as into the future (for instance to deny you a good insurance rate based on your heart rate and hospital visits). While it is hence difficult to escape the collective, societal past with all its structural inequalities, it may be just as difficult to escape one's personal past and repair one's own wrongs.[275]

---

270 Kate Crawford, *Atlas of AI* (New Haven: Yale University Press, 2021).
271 Dastin, 'Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women'.
272 Mohamed, Png, and Isaac, 'Decolonial AI'.
273 Crawford, *Atlas of AI*, 12.
274 Arendt, *The Human Condition*, 236.
275 In this regard, reference can be made to the 'right to be forgotten', acknowledged in the EU legal order to meet this problem (at least to some extent).

These three aspects, considered from the perspective of our intersubjective human condition, problematize our ability to care about the past of others as if they were our own past, and to act upon our ethical obligation to make right what was wrong.[276] Furthermore, their inhibition of our ability to fulfil this obligation also risks foreclosing us from an essential part of our – essentially ethical – human condition, and hence from that which makes our lives meaningful. Coming to terms with our past in a shared world, in which we can discuss and analyze the numerous aspects of this past, and remember that which needs to be remembered to build something better – rather than perpetuating systemic inequalities or personal mistakes – is an important part of the intersubjective reality, which risks being defied if not adequately dealt with.

(b) Present

Our experience of the present is mediated by our understanding of the past, and the anticipation of the future. We have seen how the past – along with its historical inequalities – is continuously imported into the present, under the vision of the human beings behind the AI system, rather than the views arising in a common world with a plurality of voices. At the same time, the narrative of progress that accompanies the algorithmized world, and the power that is entrenched through the use of those systems, leaves little room to contest this vision. As with historicism, also here, it is the 'victor' who decides the narrative, and hence the way in which we – in the present – look at the past and the future. In the algorithmized world, this victor is the developer of the AI system.

This can be illustrated by revisiting Walter Benjamin's *Theses on the Philosophy of History*. Interestingly, at the very beginning thereof, he confronts us with a story that deals with one of the most (in)famous historical examples of an alleged AI system or 'automaton'. This automaton – known as the Mechanical Turk – turned out to be an elaborate illusion, yet allegedly tricked people as prominent as Empress Maria Theresa of Austria and Benjamin Franklin.[277] Consider the following extract:

> The story is told of an automaton constructed in such a way that it could play a winning game of chess, answering each move of an opponent with a countermove. A puppet in Turkish attire and with a hookah in its mouth sat before a chessboard placed on a large table. A system of mirrors created the illusion that this table was transparent from all sides. Actually, a little hunchback who was an expert chess player sat inside and guided the pup pet's hand by means of strings. One can imagine a philosophical counter part to this device. The puppet called 'historical materialism' is to win all the time. It can easily be a match for anyone if it enlists the services of the ology, which today, as we know, is wizened and has to keep out of sight.[278]

We can analyze this passage on two levels: first, focusing on the myth of historical materialism, and, second, focusing on the myth of AI. The first was already discussed above. A linear

[276] See the importance thereof in Levinas, *Le Temps et l'Autre*.
[277] William Clark, Jan Golinski, and Simon Schaffer, *The Sciences in Enlightened Europe* (University of Chicago Press, 1999), 154.
[278] Benjamin, *Theses on the Philosophy of History*.

perspective of historical progress risks overlooking the debris that is caused along the way, since its narrative is always written by the victor who focuses on accomplishments rather than on the debris. As regards the second level, we need not even take recourse to a "*philosophical counter part to this device*" as Benjamin suggests, since it is the device itself that is of interest to our inquiry.

Even if AI systems today are sufficiently advanced to truly beat world champions of chess without necessitating a hunchback inside the machine, nevertheless, in one form or another, the hunchback is still there. After all, AI systems *always* rely on human beings. The idea of AI systems' objectivity or their ability to provide 'positive' rather than 'normative' outcomes merely pushes such human reliance out of sight, but does not undo it. Hence, the puppet behind the AI system – namely its creators or deployers – can "win all the time", since they are the ones deciding the function for which the system is optimized, the datasets it will be trained on, and the purposes for which it will be used. In so doing, they also decide the lens through which we look at the past and consider the future. They are hence the equivalent of the *victor* under historical materialism, at the cost of the *loser*. In the algorithmized world, the losers – individuals subjected to the AI systems, and especially those who are already in a vulnerable position – hence risk losing three times: first, when they are excluded from choosing the parameters of the system; second, when they suffer its potentially adverse impact; and, third, when they are confronted with a subsequent narrative that emphasizes the systems' beneficial role in advancing humanity's progress.

As noted above, at the heart of the issue lies the fact that these elements, while having a public impact, are not subjected to a plurality of views – including the views of those subjected thereto. This runs counter to our obligation of care towards the other, and the ideal of shaping society through collective action. Contesting the adverse impact of present AI systems, in essence, equals contesting the power of the human beings behind the system. While an individual's present condition can instantaneously be altered by the subjection to an AI system, the contestation of this impact – and the difficulty associated with contesting it[279] – instead occurs so slowly that it can make time stand still.[280] The numerous steps that typically need to be undertaken to challenge the adverse effects of the system – and the anonymity of those who are responsible for it – risk stretching the moment of contestation into what seems like eternity, until one gives up and accepts the status quo.

(c) Future

The algorithmized world also alters our experience of the future. The intersubjective time, which comes to the individual from outside[281], as an exteriority, is an essential part of the human

[279] Bart van der Sloot and Sascha van Schendel, 'Procedural Law for the Data-Driven Society', *Information & Communications Technology Law* (20 January 2021): 1–29.

[280] Compare this to the monotonous and endlessly stretching of the 'now' for Kafka's protagonist Josef K in Franz Kafka, *The Trial*, trans. Breon Mitchell (Schocken, 1999). For a philosophical discussion of the notion of time in Kafka's novel, in light of the foreclosure of the intersubjective relationship, see Luc Anckaert, 'Before the Law. Beyond Subjectivity and Objectivity', *Bruns - Europa Forum Philosophie* 14 (1999): 55–58.

[281] Anckaert, 'Before the Law. Beyond Subjectivity and Objectivity'.

condition. For Rosenzweig, it enables revelation.[282] For Levinas, it enables the meaningful life we have through the relationship with the face of the other.[283] Levinas also considers the future as an alterity, a type of *You* that is inherently different from and unknown to us. The future is indefinite and open-ended, which correlates with our human freedom.[284] Yet we deploy AI systems precisely to make models and predictions about the future, in order to understand as well as to try to shape it to our hand. In doing so, the aim is to determine the future rather than consider it as an unknown alterity. What, then, does the determination of the future through predictive AI systems – through which we turn this future *You* into an *It* – imply?

In *Le Temps et l'Autre*, Levinas cautions us against an over-objectivization of the future. "*When one deprives the present of all anticipation, the future loses all co-naturalness with it. The future is not buried in the bowels of a pre-existent eternity, where we would come to lay hold of it. It is absolutely other and new. It is thus that one can understand the very reality of time, the absolute impossibility of finding in the present the equivalent of the future, the lack of any hold upon the future*."[285] Yet with the predictive models that AI systems generate – whether it concerns the prediction of future events or our future behavior – we are doing precisely that: seeking in the present, based on past data, an equivalent in the future. As we saw above, the model will necessarily be predicated by data of the past, and hence can never entirely present us with something "absolutely other and new". The risk is that this pre-determined nature of the model is overlooked, along with the model's probabilistic nature, and that the model's outcomes are taken as reality. This, in turn, will make us act upon the predictions, thereby turning them into a self-fulfilling prophecy, and strengthening the feedback loop of the system.[286]

Accordingly, it does not matter that the model cannot truly capture the open-ended future, since by following its course, we will seek to bend the future towards its outcome. This deterministic approach to the future also has consequences upon our margin of human freedom. Our desire to know and control the future, is at the same time a rejection of the unpredictability of freedom – as entwined with Arendt's conceptualization of human action. Indeed, human being's capacity for action is not only linked to their ability to participate in the political life, but also reflects their freedom, in the form of unpredictability. Through action – and the interconnected speech

---

282 Anckaert, *God, Wereld en Mens*, 130.

283 Morgan, *The Cambridge Introduction to Emmanuel Levinas*, 172.

284 See Emmanuel Levinas, *Le Temps et l'Autre*, 11th ed. (Paris: Presses Universitaires de France (2014), 1979).

285 Translation from Morgan, *The Cambridge Introduction to Emmanuel Levinas*.

286 This can be illustrated by considering an AI-enabled predictive policing system. Police forces often have limited resources, and need to prioritize certain tasks over others in light of these limitations. Based on an analysis of past data, AI systems are deployed to map which areas of a city is most likely to be plagued by new crimes, and hence where the police should prioritize its resources. It is, however, possible that this data is biased to start with, for instance because it only entails data from some areas and not others, or it excludes data regarding white collar crimes, or it reflects information from patrols carried out by racist police officers, who patrolled more frequently in colored neighborhoods. Either way, when the AI system suggests an area that police forces should prioritize, the mere fact that those forces patrol there will result in more positive cases: where one does not look, one cannot find. This positive feedback will be provided to the AI system, which will be strengthened in its conclusion that those were the right neighborhoods to patrol, thereby merely reinforcing a problematic feedback loop rather than reflecting the diversity that makes up reality. See in this regard also O'Neil, *Weapons of Math Destruction*.

– human beings essentially disclose themselves.[287] However, human beings can impossibly predict the open-ended consequences of their actions in advance, including that which they disclose of themselves. As Arendt specifies: "*This is not simply a question of inability to foretell all the logical consequences of a particular act, in which case an electronic computer would be able to foretell the future, but arises directly out of the story which, as a result of action, begins and establishes itself as soon as the fleeting moment of the deed is past*.[288]

In other words, the activity of action as carried out with other human beings, which – contrary to work – does not fulfil a purpose of utility, is inherently unpredictable. This unpredictability, however, makes us uneasy, since it prevents us from controlling our world. We therefore try to seek certainty, and AI systems – as tools that can help us analyze the past to model and predict the future – can help us in this endeavor. Yet by substituting the unpredictability and frailty that accompanies human freedom, with certainty and reliability, we risk substituting acting with making, and end up with utility rather than meaning. Like the abovementioned artist that carves out a statue, algorithms can help us achieve a tangible product with a clearly recognizable end and greater reliability. Yet Arendt's writings caution us that this remedy can destroy the very substance of human relationships.[289] By disregarding human action and relationships in search for predictability, we risk losing that which makes life meaningful in the first place.[290]

In sum, our approach to the future in the algorithmized world – as well as to the past and present – brings under heavy strain our intersubjectivity, and imports the totalizing logic of AI also to our temporal experience. Moreover, given the inextricable link between our temporality on the one hand, and the way we engage with *alterity* and build out an intersubjective rationality on the other hand, this strain runs as a thread through AI's impact on the human condition more broadly. Indeed, we can connect the totalizing impact of AI's ubiquity on our way of experiencing history, with the totalizing tensions raised by an excessive reliance on algorithmic rationality – which reduces plurality to binarity, systematizes Goodness, and undermines our ability to engage in human relationships. Furthermore, these tendencies are likewise connected to the risks generated by the deployment of AI in the (once) public realm – including the polarization, isolation and dehumanization of individuals – and the risk of banalizing problematic or immoral decision-making rather than instilling a sense of responsibility for others. Given all of these concerns, the question that now inevitably arises is: what can we do about them? The space limitations of this paper do not allow for an extensive discussion of this question. However, based on the above, a number of conclusions can be drawn as regards potential pathways to explore – which can at the same time be read as a future research agenda.

287 Arendt, *The Human Condition*, 192.
288 Arendt, *The Human Condition*, 191.
289 Arendt, *The Human Condition*, 195.
290 Arendt, *The Human Condition*, 222.

6. CONCLUSIONS

In this paper, after describing the paradigm of the algorithmized world, and setting out how much of current ethics discourse falls short of addressing the profound impact of this paradigm on the human condition, I looked for an Archimedean meta-technological perspective that would allow me to provide a more fundamental critique on AI's ubiquity – which I found in the concept of human intersubjectivity. Through this perspective, and drawing on the insights of Jewish thinkers who emphasized the importance of relationality and its role in countering totalizing systems, I examined what it means to be human in an algorithmized world, and structured this examination around three axes: rationality, alterity and history.

My analysis is by no means exhaustive.[291] Yet by examining how the way we think, engage with others and experience time is altered by the wide-spread use of AI systems, a number of worrying trends came to the surface, which evoked the same risks that can be encountered in totalitarian visions. It goes beyond the purpose of this paper to provide a comprehensive overview of the measures that should be taken to counter these worrying trends. It even goes beyond the purpose of this paper to examine whether the tipping point has not already been reached, and to which extent 'countering' is still an option. Yet to conclude this paper, I nevertheless briefly consider how the identified challenges could be addressed, assuming that past insights on the importance of human relationships are still relevant today.

6.1 Acknowledging the extent of the problem

As with all problems, the first step towards a solution consists in its acknowledgment. While the acknowledgment that AI brings forth ethical – as well as legal, social and other – risks is no longer at stake (as Chapter 4.1 has shown, this is by now well-established), the acknowledgment of the more fundamental, human condition-altering impact of AI, is still less evident. Current ethics discourse is tolerant of ethics guidelines, and even of the translation of such guidelines into binding legislation, yet if we recall the profoundness of the risks associated with the banalization of ethically problematic decisions by means of AI, and the fact that our own human psyche has inbuilt mechanisms to evade responsibility – an evasion that is only strengthened by the use of AI and its opacity – it is difficult to maintain that the current steps are sufficient. To put it bluntly: Eichmann would probably not have been served with a set of ethics guidelines. Without denying the need for awareness-raising, education programs, guidelines, and most particularly, binding legal rules, there is an equally urgent need for a more fundamental analysis and critique of the issues at stake, which requires us to look at the use of these systems with a different attitude altogether – and dare to let go of the narrative of progress.

The questions that we need to ask ourselves – individually and collectively – are not only what the minimum ethical requirements are that AI systems should comply with. They should focus on how we can mitigate the totalizing tendencies that AI systems are exacerbating, and what we

[291] Furthermore, I did not cover the ways in which human intersubjectivity can be positively be impacted by AI systems – a subject that is also valid yet goes beyond the purpose of this inquiry, and that all too often risks inviting a utilitarian cost-benefit analysis, which is the opposite of my aim here.

can do to counter the consequences thereof for society at large, acknowledging that this concerns societal harm rather than mere individual and collective harm. Moreover, our questions should focus on how we can deal with the loss of human responsibility and accountability – consciously or unconsciously, deliberately or unwillingly – in a world where the delegation of authority to self-learning machines is becoming a normality. This means we need to rethink the importance of enabling Levinas' face-to-face encounter, and strengthening human relationships to ensure that the ethical obligations we have towards others are not undermined by layers of code. This also means, as Arendt advocated, that the scientific world should be reunited with the political.[292] If we know that a precondition for scientists' sense of responsibility (like AI developers), consists in the ability to adequately judge situations – in light of not only internal consistency, but also a verification with common sense, within a 'common world' – we need to find pathways to restore and maintain that common world.

6.2 Carving out spaces for action

That common world is on shaky ground, and certainly not only due to the ubiquity of AI. For a number of reasons, popularism is on the rise,[293] and it thrives on the polarization of society. As we examined above, the irresponsible deployment of AI systems – in particular on social media platforms, which have become a treasured political arena, but also in public spaces more generally – can exacerbate this polarizing tendency, while at the same time atomizing individuals to undermine their ability to organize and challenge the unjust exercise of power. We therefore need to examine how, instead, the broken bonds between human beings can be mended, across political, social and generational boundaries.[294] This mending process is not only important to enable a pluralistic deliberation process through which we can discuss the major challenges of our time, but also to address the abovementioned problems of AI. As has been clarified through the above analysis, the core of the problem is not technical, but social and political.[295] Yet by excessively focusing on the technical aspects, we risk forgetting this bigger picture, in which accumulative problems are slowly bringing to boil the water in which we are bathing.

---

292 See also the Foreword by Danielle Allen, xv, in Arendt, *The Human Condition*.

293 Michael Cox, 'Understanding the Global Rise of Populism', Strategic Update (London: London School of Economics, February 2018); Thomas M. Meyer and Markus Wagner, 'The Rise of Populism in Modern Democracies', in *The Oxford Handbook of Political Representation in Liberal Democracies*, ed. Robert Rohrschneider and Jacques Thomassen (Oxford University Press, 2020), 562–81; Pippa Norris, 'The Populist Challenge to Liberal Democracies', in *The Oxford Handbook of Political Representation in Liberal Democracies*, ed. Robert Rohrschneider and Jacques Thomassen (Oxford University Press, 2020), 543–62.

294 Verovšek, 'Integration after Totalitarianism', 8.

295 This point is well-made by, inter alia, Zuboff, *The Age of Surveillance Capitalism*; Gry Hasselbalch, 'Making Sense of Data Ethics. The Powers behind the Data Ethics Debate in European Policymaking', *Internet Policy Review* 8, no. 2 (13 June 2019); Julie E. Cohen, *Between Truth and Power: The Legal Constructions of Informational Capitalism* (New York, NY: Oxford University Press, 2019); Pratyusha Kalluri, 'Don't Ask If Artificial Intelligence Is Good or Fair, Ask How It Shifts Power', *Nature* 583, no. 7815 (7 July 2020): 169–169; Crawford, *Atlas of AI*.

By carving out spaces where human beings can come together, act and speak[296], in defiance of statistical laws and probabilistic models[297], and in recognition of the human frailty that is part of their uniqueness, we could tackle multiple issues at once. First, we could delineate domains in which the mathematical functions of AI systems – and the algorithmic reductionist paradigm that underpin them – have no place. Second, we could strengthen the plurality of voices in the public sphere and build a stronger foundation for the 'common sense' that enables critical thinking and adequate judgment, to counter the technocratic rationality and the AI systems that give expression to it.[298] Third, we could subject the value-laden and normative choices that underlay the development and deployment of AI systems to democratic oversight, informed by an open and public debate, which includes the individuals subjected to AI systems. Rather than being reduced to an *It* in the process, the irreducibility and alterity of these individuals should be respected, also by those who develop and deploy AI systems.

6.3 Combatting binarity

A third pathway to explore, concerns the way in which we can harness that which falls outside of a rational binarity, and give that which is out of place, a place in the algorithmized world. There is an important difference between acknowledging the need for a plurality of voices in order to enable and maintain a common world and space for meaningful human action, and adopting an organizational policy that is meant to tick some diversity boxes. As is the case for ticking the boxes of ethical checklists for AI, they can treat a symptom, but not provide a cure. Harnessing plurality, and opening up the possibility for deviations, declinations and statistical outliers – or Deleuze's 'case vide'[299] – requires a change of mindset: one in which the meaning that can arise therefrom is appreciated rather than ignored for not falling within the norm.

Once again, this means identifying the domains in which a reductionist approach to the human condition should be shunned, since – artificial – predictability and control comes at too great of a cost. It also means leaving space for the phenomenon of the little goodness, which we described above. Despite insistence on face-to-face encounters, Levinas ultimately also acknowledges that the presence of *many* others – including those who are not yet born – eventually requires a context in which responsibility can be organized in a structural way.[300] Yet if systems are inescapable, we must at least continuously and critically evaluate them, and take timely action when they fail to ensure justice for those many others – especially outliers – in a way that respects their human dignity. While we can link this need to the protection offered by

---

296 See in this regard also Dan McQuillan, 'The Political Affinities of AI', in *The Democratization of Artificial Intelligence*, ed. Andreas Sudmann (transcript Verlag, 2019), 163–74.

297 Arendt, *The Human Condition*, 178.

298 We can recall Stanley Milgram's experiment, and note that the introduction of a dissenting voice as part of the experiment reduced the number of people who administered further shocks. This can be taken as a confirmation of the importance of ensuring free speech, democratic participation and pluralism. See also Hugh Murray, review of *Review of Modernity and the Holocaust*, by Zygmunt Bauman, *German Politics & Society*, no. 22 (1991): 86.

299 See in this regard also Dufour, *Les mystères de la trinité*, 31; Évelyne Grossman, 'Structuralisme et métaphysique', *Litterature* n°167, no. 3 (4 October 2012): 129.

300 See for instance Lévinas, *Is It Righteous to Be?* See in this regard also Luc Anckaert, 'Ethics of Responsibility and Ambiguity of Politics in Levinas's Philosophy', *Problemos* 97 (21 April 2020): 70.

a robust legal system of human rights, there is still work to do to translate those rights to the context of AI, and to ensure a broader societal environment in which those rights can be enabled in the first place, including through democracy and the rule of law.[301]

Ultimately, AI systems are but technical tools, within a much broader fabric that is made up by our society, culture, economy, political institutions and laws – all of which interact together and make up the world we live in. Nevertheless, the totalizing logos of those systems – even when driven by a desire to improve human lives – risks undermining the very essence of what makes human life meaningful. Taking a meta-technological perspective can help us to shed light on these risks and to provide a critique that goes well beyond current ethics discourse, which is running against its limits and requires a broadening of perspective. There remains, however, still a lot of work to map – ideally from a multidisciplinary perspective – the various ways in which the algorithmized world is impacting and altering the intersubjective nature of the human condition. The purpose of this paper is to contribute to this mapping work and to demonstrate that the insights of twentieth-century Jewish thinkers – by addressing the risks of totalizing systems and emphasizing the value of human relationships – can provide a welcome starting point to do so.

[301] Smuha, 'Beyond a Human Rights-Based Approach to AI Governance'.

BIBLIOGRAPHY

Adorno, Theodor W. *Negative Dialektik. Jargon der Eigentlichkeit*. Frankfurt am Main: Suhrkamp, 1973.

Ala-Pietilä, Pekka, and Nathalie A. Smuha. 'A Framework for Global Cooperation on Artificial Intelligence and Its Governance'. In *Reflections on Artificial Intelligence for Humanity*, edited by Bertrand Braunschweig and Malik Ghallab, 237–65. Cham: Springer International Publishing, 2021. https://doi.org/10.1007/978-3-030-69128-8_15.

Alexander, Lawrence. 'Scalar Properties, Binary Judgments'. *Legal Studies Research Paper Series*, no. Research Paper No. 07-19 (October 2005).

AlgorithmWatch. 'Automating Society Report 2020', October 2020. https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/10/Automating-Society-Report-2020.pdf.

Allen, Amy. *The End of Progress: Decolonizing the Normative Foundations of Critical Theory*. New York: Columbia University Press, 2016.

Alwosheel, Ahmad, Sander van Cranenburgh, and Caspar G. Chorus. '"Computer Says No" Is Not Enough: Using Prototypical Examples to Diagnose Artificial Neural Networks for Discrete Choice Analysis'. *Journal of Choice Modelling* 33 (1 December 2019): 100186. https://doi.org/10.1016/j.jocm.2019.100186.

Ananny, Mike. 'Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness'. *Science, Technology, & Human Values* 41, no. 1 (January 2016): 93–117. https://doi.org/10.1177/0162243915606523.

Anckaert, Luc. *A Critique of Infinity: Rosenzweig and Levinas*. Studies in Philosophical Theology 35. Leuven: Peeters, 2006.

———. 'Before the Law. Beyond Subjectivity and Objectivity'. *Bruns - Europa Forum Philosophie* 14 (1999): 55–58.

———. 'Ethics of Responsibility and Ambiguity of Politics in Levinas's Philosophy'. *Problemos* 97 (21 April 2020): 61–74. https://doi.org/10.15388/Problemos.97.5.

———. 'Franz Rosenzweigs Stern der Erlösung. Een hermeneutische en retorische benadering'. In *Joodse filosofie tussen rede en traditie. Feestbundel ter ere van de tachtigste verjaardag van Prof. dr. H.J. Heering*, 223–41. Kampen: Uitgeverij Kok, 1993.

———. 'Globalisation and the Tragedy of Ethics'. In *Building Towers: Perspectives on Globalisation*, edited by Luc Anckaert, Danny Cassimon, and Hendrik Opdebeeck, 9–36. Ethical Perspectives Monograph Series 2. Leuven: Peeters, 2002.

———. *God, wereld en mens: het ternaire denken van Franz Rosenzweig*. Wijsgerige verkenningen 17. Leuven: Universitaire Pers Leuven, 1997.

———. 'Goodness without Witnesses: Vasily Grossman and Emmanuel Levinas'. In *Levinas and Literature*, edited by Michael Fagenblat and Arthur Cools, 223–38. De Gruyter, 2020. https://doi.org/10.1515/9783110668926-015.

———. 'Language, Ethics, and the Other between Athens and Jerusalem. A Comparative Study of Plato and Rosenzweig'. *Philosophy East and West* 45, no. 4 (1 January 1995): 545–67.

———. 'The Thunderbolt of Evil and Goodness without Witnesses: In Conversation with Vasili Grossman, Life and Fate'. *Religija Ir Kultūra* 18–19 (2016): 22–37.

Andersen, Niklas Andreas. 'The Technocratic Rationality of Governance - the Case of the Danish Employment Services'. *Critical Policy Studies*, 28 December 2020, 1–19. https://doi.org/10.1080/19460171.2020.1866629.

Arendt, Hannah. *Eichmann in Jerusalem: A Report on the Banality of Evil*. Viking Press, 1963.

———. *Lectures on Kant's Political Philosophy*. Chicago: University of Chicago Press, 1982.

———. *The Human Condition*. University of Chicago Press (2019), 1958.

———. *The Origins of Totalitarianism*. Penguin Classics (2017), 1951.

Bahner, J. Elin, Anke-Dorothea Hüper, and Dietrich Manzey. 'Misuse of Automated Decision Aids: Complacency, Automation Bias and the Impact of Training Experience'. *International Journal of Human-Computer Studies* 66, no. 9 (September 2008): 688–99. https://doi.org/10.1016/j.ijhcs.2008.06.001.

Bauman, Zygmunt. *Modernity and Ambivalence*. Cambridge: Polity Press, 1991.

———. *Modernity and the Holocaust*. Cambridge: Polity press, 1989.

Baxter, Gordon, and Ian Sommerville. 'Socio-Technical Systems: From Design Methods to Systems Engineering'. *Interacting with Computers* 23, no. 1 (1 January 2011): 4–17. https://doi.org/10.1016/j.intcom.2010.07.003.

Bayamlıoğlu, Emre, and Ronald Leenes. 'The "Rule of Law" Implications of Data-Driven Decision-Making: A Techno-Regulatory Perspective'. *Law, Innovation and Technology* 10, no. 2: 295–313. https://doi.org/10.1080/17579961.2018.1527475.

Bedingfield, Will. 'Everything That Went Wrong with the Botched A-Levels Algorithm'. *Wired UK*, 19 August 2020. https://www.wired.co.uk/article/alevel-exam-algorithm.

Beiner, Ronald. 'Walter Benjamin's Philosophy of History'. *Political Theory* 12, no. 3 (1984): 423–34.

Benjamin, Ruha. *Race After Technology: Abolitionist Tools for the New Jim Code*. 1 edition. Medford, MA: Polity, 2019.

Benjamin, Walter. *Theses on the Philosophy of History*. New York: Schocken Books, 1968.

Benzécri, J.-P. *L'analyse des données. 2: L'analyse des correspondances*. Leçons sur l'analyse factorielle et la reconnaissance des formes et travaux du Laboratoire de statistique de l'Université de Paris VI. Paris: Dunod, 1973.

Berendt, Bettina. 'AI for the Common Good?! Pitfalls, Challenges, and Ethics Pen-Testing'. *Paladyn, Journal of Behavioral Robotics* 10, no. 1: 44–65. https://doi.org/10.1515/pjbr-2019-0004.

Bietti, Elettra. 'From Ethics Washing to Ethics Bashing'. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, January 2020, 210–19. https://doi.org/10.1145/3351095.3372860.

Binns, Reuben. 'On the Apparent Conflict Between Individual and Group Fairness'. *ArXiv:1912.06883 [Cs, Stat]*, 14 December 2019. http://arxiv.org/abs/1912.06883.

Black, Max. 'The Gap Between "Is" and "Should"'. *The Philosophical Review* 73, no. 2 (1964): 165–81. https://doi.org/10.2307/2183334.

Boddington, Paula. *Towards a Code of Ethics for Artificial Intelligence*. Artificial Intelligence: Foundations, Theory, and Algorithms. Cham: Springer International Publishing, 2017. https://doi.org/10.1007/978-3-319-60648-4.

Boutyline, Andrei, and Robb Willer. 'The Social Structure of Political Echo Chambers: Variation in Ideological Homophily in Online Networks'. *Political Psychology* 38, no. 3 (2017): 551–69.

Bowker, Geoffrey C., and Susan Leigh Star. 'Building Information Infrastructures for Social Worlds — The Role of Classifications and Standards'. In *Community Computing and Support Systems: Social Interaction in Networked Communities*, edited by Toru Ishida,

231–48. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 1998. https://doi.org/10.1007/3-540-49247-X_16.

Boym, Svetlana. 'From Love to Worldliness: Hannah Arendt and Martin Heidegger'. *The Yearbook of Comparative Literature* 55, no. 1 (2009): 106–28. https://doi.org/10.1353/cgl.2011.0003.

Brkan, M. 'Artificial Intelligence and Democracy'. *Delphi - Interdisciplinary Review of Emerging Technologies* 2, no. 2 (2019): 66–71. https://doi.org/10.21552/delphi/2019/2/4.

Brownsword, Roger. 'Technological Management and the Rule of Law'. *Law, Innovation and Technology* 8, no. 1: 100–140. https://doi.org/10.1080/17579961.2016.1161891.

Brundage, Miles, and et al. 'The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation', February 2018. https://maliciousaireport.com/.

Buber, Martin. *I and Thou*. Translated by Walter Kaufman. New York: Simon and Schuster (2000), 1923.

Buckler, Steve. *Hannah Arendt and Political Theory: Challenging the Tradition*. Edinburgh University Press, 2011.

Buiten, Miriam C. 'Towards Intelligent Regulation of Artificial Intelligence'. *European Journal of Risk Regulation* 10, no. 1 (March 2019): 41–59. https://doi.org/10.1017/err.2019.8.

Buolamwini, Joy, and Timnit Gebru. 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification'. In *Proceedings of Machine Learning Research*, 81:1–15, 2018. http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf.

Cadwalladr, Carole. 'Fresh Cambridge Analytica Leak "Shows Global Manipulation Is out of Control"'. *The Guardian*, 4 January 2020, sec. UK news. http://www.theguardian.com/uk-news/2020/jan/04/cambridge-analytica-data-leak-global-election-manipulation.

Cath, Corinne, Sandra Wachter, Brent Mittelstadt, Mariarosaria Taddeo, and Luciano Floridi. 'Artificial Intelligence and the "Good Society": The US, EU, and UK Approach'. *Science and Engineering Ethics*, 28 March 2017. https://doi.org/10.1007/s11948-017-9901-7.

Chowdhury, Rumman. 'Sharing Learnings about Our Image Cropping Algorithm'. Twitter, 19 May 2021. https://blog.twitter.com/engineering/en_us/topics/insights/2021/sharing-learnings-about-our-image-cropping-algorithm.

Churchland, Patricia Smith. *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. 2nd print. Cambridge (Mass.): MIT press, 1986.

Clark, William, Jan Golinski, and Simon Schaffer. *The Sciences in Enlightened Europe*. University of Chicago Press, 1999.

Coeckelbergh, Mark. *AI Ethics*. The MIT Press Essential Knowledge Series. Cambridge, Mass: MIT Press, 2020.

———. 'When Machines Talk: A Brief Analysis of Some Relations between Technology and Language'. *Technology and Language* 1, no. 1 (2020): 22–27. https://doi.org/10.48417/TECHNOLANG.2020.01.05.

Cohen, Julie E. *Between Truth and Power: The Legal Constructions of Informational Capitalism*. New York, NY: Oxford University Press, 2019.

Cole, David. 'The Chinese Room Argument'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2020. Metaphysics Research Lab, Stanford University, 2020. https://plato.stanford.edu/archives/win2020/entries/chinese-room/.

Connelly, James. ‘Facing the Past: Walter Benjamin’s Antitheses’. *The European Legacy* 9, no. 3 (June 2004): 317–29. https://doi.org/10.1080/1084877042000235487.

Coughlan, Sean. ‘Why Did the A-Level Algorithm Say No?’ *BBC News*, 14 August 2020, sec. Family & Education. https://www.bbc.com/news/education-53787203.

Council of Europe Ad Hoc Committee on Artificial Intelligence (CAHAI). ‘Feasibility Study’. Strasbourg: Council of Europe, 17 December 2020. https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da.

Cowls, Josh, Thomas King, Mariarosaria Taddeo, and Luciano Floridi. ‘Designing AI for Social Good: Seven Essential Factors’. *SSRN Electronic Journal*, 2019. https://doi.org/10.2139/ssrn.3388669.

Cox, Michael. ‘Understanding the Global Rise of Populism’. Strategic Update. London: London School of Economics, February 2018. https://www.lse.ac.uk/ideas/Assets/Documents/updates/LSE-IDEAS-Understanding-Global-Rise-of-Populism.pdf.

Crawford, Kate. *Atlas of AI*. New Haven: Yale University Press, 2021.

Crawford, Kate, Roel Dobbe, Theodora Dryer, Genevieve Fried, Ben Green, Elizabeth Kaziunas, Amba Kak, et al. *AI Now 2019 Report*. New York: AI Now Institute, 2019. https://ainowinstitute.org/AI_Now_2019_Report.pdf.

Crawford, Kate, and Meredith Whittaker. ‘The AI Now Report - The Social and Economic Implications of Artificial Intelligence Technologies in the Near Term’. New York: The AI Now Institute, 2016. https://ainowinstitute.org/AI_Now_2016_Report.pdf.

Crowell, Steven. ‘Why Is Ethics First Philosophy? Levinas in Phenomenological Context’. *European Journal of Philosophy* 23, no. 3 (2015): 564–88. https://doi.org/10.1111/j.1468-0378.2012.00550.x.

Cummings, Mary. ‘Automation Bias in Intelligent Time Critical Decision Support Systems’. *American Institute of Aeronautics and Astronautics*, 1 November 2014. https://web.archive.org/web/20141101113133/http://web.mit.edu/aeroastro/labs/halab/papers/CummingsAIAAbias.pdf.

Dastin, Jeffrey. ‘Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women’. *Reuters*. 10 October 2018, sec. Retail. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.

Deloitte. ‘Bringing transparency and ethics in AI’. Deloitte Netherlands. Accessed 12 August 2021. https://www2.deloitte.com/nl/nl/pages/innovatie/artikelen/bringing-transparency-and-ethics-into-ai.html.

Diamond, Larry. ‘The Threat of Postmodern Totalitarianism’. *Journal of Democracy* 30, no. 1 (2019): 20–24. https://doi.org/10.1353/jod.2019.0001.

Dignum, Virginia. *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Artificial Intelligence: Foundations, Theory, and Algorithms. Springer International Publishing, 2019.

Ding, Han, Robert X. Gao, Alf J. Isaksson, Robert G. Landers, Thomas Parsini, and Ye Yuan. ‘State of AI-Based Monitoring in Smart Manufacturing and Introduction to Focused Section’. *IEEE/ASME Transactions on Mechatronics* 25, no. 5 (2020): 2143–54. https://doi.org/doi: 10.1109/TMECH.2020.3022983.

Diver, Laurence. ‘Interpreting the Rule(s) of Code: Performance, Performativity, and Production’. *MIT Computational Law Report*, 15 July 2021. https://law.mit.edu/pub/interpretingtherulesofcode/release/1.

Dufour, Dany-Robert. *Les mystères de la trinité*. Bibliothèque des sciences humaines. Paris: Gallimard, 1990.

Durkheim, Emile. *L'éducation Morale*. Paris: Alcan, 1925.

European Commission. 'Artificial Intelligence for Europe', 25 April 2018. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51625.

———. Proposal for a Regulation of the European Parliament and the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts., Pub. L. No. COM(2021) 206 final, 2021/0106 (COD) (2021).

———. 'White Paper On Artificial Intelligence - A European Approach to Excellence and Trust'. Brussels, 19.2.2020 COM(2020) 65 final, 19 February 2020.

European Parliament and Council. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), § OJ L 119 (2016). https://eur-lex.europa.eu/eli/reg/2016/679/oj.

Feinberg, Joel. 'Harm to Others'. In *The Moral Limits of the Criminal Law - Volume 1: Harm to Others*. New York: Oxford University Press, 1984.

Flanagan, Kieran. 'Bauman's Travels: Metaphors of the Token and the Wilderness'. In *Liquid Sociology: Metaphor in Zygmunt Bauman's Analysis of Modernity*, edited by Mark Davis. Routledge, 2016.

Flaxman, Seth, Sharad Goel and Justin M. Rao, 'Filter Bubbles, Echo Chambers, and Online News Consumption', *Public Opinion Quarterly* 80, no. S1 (1 January 2016): 298–320, https://doi.org/10.1093/poq/nfw006.

Floridi, Luciano. 'On Human Dignity as a Foundation for the Right to Privacy'. *Philosophy & Technology* 29, no. 4 (December 2016): 307–12. https://doi.org/10.1007/s13347-016-0220-8.

———. 'The Ontological Interpretation of Informational Privacy'. *Ethics and Information Technology* 7, no. 4 (December 2005): 185–200. https://doi.org/10.1007/s10676-006-0001-7.

Gebru, Timnit. 'Race and Gender'. In *The Oxford Handbook of Ethics of AI*, by Timnit Gebru, 251–69. edited by Markus D. Dubber, Frank Pasquale, and Sunit Das. Oxford University Press, 2020. https://doi.org/10.1093/oxfordhb/9780190067397.013.16.

Gellers, Joshua C. *Rights for Robots: Artificial Intelligence, Animal and Environmental Law*. Routledge, 2020.

Goddard, Kate, Abdul Roudsari, and Jeremy C. Wyatt. 'Automation Bias: Empirical Results Assessing Influencing Factors'. *International Journal of Medical Informatics* 83, no. 5 (1 May 2014): 368–75. https://doi.org/10.1016/j.ijmedinf.2014.01.001.

Goldberg, David Theo. 'Coding Time'. *Critical Times* 2, no. 3 (1 December 2019): 353–69. https://doi.org/10.1215/26410478-7862517.

Google. 'AI at Google: Our Principles'. Google AI. Accessed 12 August 2021. https://ai.google/principles/.

Greene, Daniel, Anna Lauren Hoffmann, and Luke Stark. 'Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning'. *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.

Griffiths, Thomas L. ‘Understanding Human Intelligence through Human Limitations’. *Trends in Cognitive Sciences* 24, no. 11: 873–83. https://doi.org/10.1016/j.tics.2020.09.001.

Grossman, Évelyne. ‘Structuralisme et métaphysique’. *Litterature* n°167, no. 3 (4 October 2012): 127–37.

Grossman, Vasily. *Life And Fate*. Translated by Robert Chandler. London: Vintage Classic, 2017.

Gunkel, David. *Robot Rights*. MIT Press, 2018.

Gunkel, David J. ‘The Other Question: Can and Should Robots Have Rights?’ *Ethics and Information Technology* 20, no. 2 (1 June 2018): 87–99. https://doi.org/10.1007/s10676-017-9442-4.

Hagendorff, Thilo. ‘The Ethics of AI Ethics: An Evaluation of Guidelines’. *Minds and Machines* 30, no. 1: 99–120. https://doi.org/10.1007/s11023-020-09517-8.

Handelman, Susan. ‘Facing the Other: Levinas, Perelman and Rosenzweig’. *Religion & Literature* 22, no. 2/3 (1990): 61–84.

———. ‘Walter Benjamin and the Angel of History’. *CrossCurrents* 41, no. 3 (1991): 344–52.

Hanna, Alex, Emily Denton, Andrew Smart, and Jamila Smith-Loud. ‘Towards a Critical Race Methodology in Algorithmic Fairness’. *ArXiv:1912.03593 [Cs]*, 7 December 2019. https://doi.org/10.1145/3351095.3372826.

Hänold, Stefanie. ‘Profiling and Automated Decision-Making: Legal Implications and Shortcomings’. In *Robotics, AI and the Future of Law*, edited by Marcelo Corrales, Mark Fenwick, and Nikolaus Forgó, 123–53. Perspectives in Law, Business and Innovation. Singapore: Springer, 2018. https://doi.org/10.1007/978-981-13-2874-9_6.

Hao, Karen. ‘He Got Facebook Hooked on AI. Now He Can’t Fix Its Misinformation Addiction’. MIT Technology Review, 11 March 2021. https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/.

———. ‘Training a Single AI Model Can Emit as Much Carbon as Five Cars in Their Lifetimes’. *MIT Technology Review* (blog), 6 June 2019. https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/.

Harari, Yuval Noah. ‘Who Will Win the Race for AI?’ *Foreign Policy* (blog). Accessed 15 July 2020. https://foreignpolicy.com/gt-essay/who-will-win-the-race-for-ai-united-states-china-data/.

Hasselbalch, Gry. ‘Making Sense of Data Ethics. The Powers behind the Data Ethics Debate in European Policymaking’. *Internet Policy Review* 8, no. 2. https://doi.org/10.14763/2019.2.1401.

———. *Data Ethics of Power* (Edward Elgar Publishing, 2021).

Hawley, Scott H. ‘Challenges for an Ontology of Artificial Intelligence’. *Perspectives on Science and Christian Faith* 71, no. 2 (2019): 83–95.

Heaven, Will Douglas. ‘Hundreds of AI Tools Have Been Built to Catch Covid. None of Them Helped.’ MIT Technology Review, 30 July 2021. https://www.technologyreview.com/2021/07/30/1030329/machine-learning-ai-failed-covid-hospital-diagnosis-pandemic/.

High-Level Expert Group on AI. ‘A Definition of AI: Main Capabilities and Scientific Disciplines’, 8 April 2019.

———. ‘Ethics Guidelines for Trustworthy AI’, 8 April 2019.

———. ‘Policy and Investment Recommendations for Trustworthy AI’, 26 June 2019.

Hildebrandt, Mireille. ‘Algorithmic Regulation and the Rule of Law’. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, no. 2128 (13 September 2018): 20170355. https://doi.org/10.1098/rsta.2017.0355.

Hildebrandt, Mireille, and Bert-Jaap Koops. ‘The Challenges of Ambient Law and Legal Protection in the Profiling Era’. *Modern Law Review* 73, no. 3 (2010): 428–60.

Hill, Kashmir. ‘Wrongfully Accused by an Algorithm’. *The New York Times*, 24 June 2020, sec. Technology. https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html.

Human Rights Watch. ‘China’s Algorithms of Repression: Reverse Engineering a Xinjiang Police Mass Surveillance App’. Human Rights Watch, 1 May 2019. https://www.hrw.org/report/2019/05/01/chinas-algorithms-repression/reverse-engineering-xinjiang-police-mass-surveillance.

Hume, David. *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects and Dialogues Concerning Natural Religion*. Edited by L. A. Selby-Bigge. Oxford: Clarendon Press (1896), 1739.

Husson, François, Julie Josse, and Gilbert Saporta. ‘Jan de Leeuw and the French School of Data Analysis’. *Journal of Statistical Software* 73, no. 6 (2016): 16 p. https://doi.org/10.18637/jss.v073.i06.

IBM Research. ‘Introducing AI Fairness 360, A Step Towards Trusted AI’, September 2018. https://www.ibm.com/blogs/research/2018/09/ai-fairness-360/.

Information Commissioner’s Office. ‘Guidance on AI and Data Protection’. ICO, July 2021. https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/guidance-on-ai-and-data-protection/.

Isaak, Jim, and Mina J. Hanna. ‘User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection’. *Computer* 51, no. 8 (August 2018): 56–59. https://doi.org/10.1109/MC.2018.3191268.

Jobin, Anna, Marcello Ienca, and Effy Vayena. ‘The Global Landscape of AI Ethics Guidelines’. *Nat Mach Intell* 1 (2019): 389–99.

Johnson, Khari, ‘AI Ethics Pioneer’s Exit from Google Involved Research into Risks and Inequality in Large Language Models’, VentureBeat, 3 December 2020. https://venturebeat.com/2020/12/03/ai-ethics-pioneers-exit-from-google-involved-research-into-risks-and-inequality-in-large-language-models/.

Jones, Steven E. *Against Technology: From the Luddites to Neo-Luddism*. 1st ed. New York: Routledge, 2006.

Kafka, Franz. *The Trial*. Translated by Breon Mitchell. Schocken, 1999.

Kalluri, Pratyusha. ‘Don’t Ask If Artificial Intelligence Is Good or Fair, Ask How It Shifts Power’. *Nature* 583, no. 7815 (7 July 2020): 169–169. https://doi.org/10.1038/d41586-020-02003-2.

Kamilaris, Andreas, and Francesc X. Prenafeta-Boldú. ‘Deep Learning in Agriculture: A Survey’. *Computers and Electronics in Agriculture* 147 (1 April 2018): 70–90. https://doi.org/10.1016/j.compag.2018.02.016.

Kernohan, Andrew. ‘Accumulative Harms and the Interpretation of the Harm Principle’. *Social Theory and Practice* 19, no. 1 (1993): 51–72.

Kitchens, Brent, Steven L. Johnson, and Peter Gray. ‘Understanding Echo Chambers and Filter Bubbles: The Impact of Social Media on Diversification and Partisan Shifts in News

Consumption'. *MIS Quarterly* 44, no. 4 (December 2020): 1619–49. https://doi.org/10.25300/MISQ/2020/16371.

Knight, Will. 'The Apple Card Didn't "See" Gender—and That's the Problem'. *Wired*, 2019. https://www.wired.com/story/the-apple-card-didnt-see-genderand-thats-the-problem/.

Kranzberg, Melvin. 'Technology and History: "Kranzberg's Laws"'. *Bulletin of Science, Technology & Society* 15, no. 1 (1 February 1995): 5–13. https://doi.org/10.1177/027046769501500104.

Kurzweil, Raymond. *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*. New York: Viking, 1999.

Laufer, William S. 'Social Accountability and Corporate Greenwashing'. *Journal of Business Ethics* 43, no. 3 (2003): 253–61. https://doi.org/10.1023/A:1022962719299.

Lévinas, Emmanuel. *Is It Righteous to Be?: Interviews with Emmanuel Lévinas*. Edited by Jill Robbins. Stanford University Press, 2001.

Levinas, Emmanuel. *Le Temps et l'Autre*. 11th ed. Paris: Presses Universitaires de France (2014), 1979.

———. *Totalité et Infini - Essai Sur l'exteriorité*. Paris: Le Livre de Poche (2021), 1971.

Lueg, Klarissa, and Rainer Lueg. 'Detecting Green-Washing or Substantial Organizational Communication: A Model for Testing Two-Way Interaction Between Risk and Sustainability Reporting'. *Sustainability* 12, no. 6 (January 2020): 2520. https://doi.org/10.3390/su12062520.

Lynskey, Orla. *The Foundations of EU Data Protection Law*. Oxford: Oxford University Press, 2015.

Maimonides, Moses. *The Guide for the Perplexed*. Translated by M. Friedländer. 4th ed. New York: E. P. Dutton & Company, 1904.

Martin, Justin D, Fouad Hassan, George Anghelcev, Noor Abunabaa, and Sarah Shaath. 'From Echo Chambers to "Idea Chambers": Concurrent Online Interactions with Similar and Dissimilar Others'. *International Communication Gazette*, 16 February 2021. https://doi.org/10.1177/1748048521992486.

Marx, Karl. *Differenz Der Demokritischen Und Epikureischen Naturphilosophie - Doktordissertation (1841)*. Hofenberger, 2014.

Mayer-Schönberger, Viktor, and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt, 2013.

McCarthy, John, M. L. Minsky, N. Rochester, and C. E. Shannon. 'A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence', 31 August 1955. http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf.

McKinsey. 'Notes from the AI Frontier: Modeling the Impact of AI on the World Economy'. Discussion Paper. McKinsey, September 2018. https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20frontier%20Modeling%20the%20impact%20of%20AI%20on%20the%20world%20economy/MGI-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-economy-September-2018.pdf.

McQuillan, Dan. 'The Political Affinities of AI'. In *The Democratization of Artificial Intelligence*, edited by Andreas Sudmann, 163–74. transcript Verlag, 2019. https://doi.org/10.14361/9783839447192-010.

Merry, Sally Engle. 'Measuring the World: Indicators, Human Rights, and Global Governance'. *Current Anthropology* 52, no. S3 (April 2011): S83–95. https://doi.org/10.1086/657241.

Meyer, Thomas M., and Markus Wagner. 'The Rise of Populism in Modern Democracies'. In *The Oxford Handbook of Political Representation in Liberal Democracies*, edited by Robert Rohrschneider and Jacques Thomassen, 562–81. Oxford University Press, 2020. https://doi.org/10.1093/oxfordhb/9780198825081.013.29.

Milgram, Stanley. *Obedience to Authority: An Experimental View*. New York: Harper Perennial, 2009.

Minardi, Di. 'The Grim Fate That Could Be "Worse than Extinction"'. *BBC*, 16 October 2020. https://www.bbc.com/future/article/20201014-totalitarian-world-in-chains-artificial-intelligence.

Mittelstadt, Brent. 'Principles Alone Cannot Guarantee Ethical AI'. *Nature Machine Intelligence* 1, no. 11 (November 2019): 501–7. https://doi.org/10.1038/s42256-019-0114-4.

Mohamed, Shakir, Marie-Therese Png, and William Isaac. 'Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence'. *Philosophy & Technology*, 12 July 2020. https://doi.org/10.1007/s13347-020-00405-8.

Mohanty, Priya. 'Do You Fear Artificial Intelligence Will Take Your Job?' Forbes, 6 July 2018. https://www.forbes.com/sites/theyec/2018/07/06/do-you-fear-artificial-intelligence-will-take-your-job/.

Morgan, Michael L. *Discovering Levinas*. Cambridge: Cambridge University Press, 2007.

———. *The Cambridge Introduction to Emmanuel Levinas*. Cambridge: Cambridge University Press, 2011.

Morley, Jessica, Anat Elhalal, Francesca Garcia, Libby Kinsey, Jakob Mökander, and Luciano Floridi. 'Ethics as a Service: A Pragmatic Operationalisation of AI Ethics'. *Minds and Machines* 31, no. 2: 239–56. https://doi.org/10.1007/s11023-021-09563-w.

Muller, Catelijne. 'The Impact of Artificial Intelligence on Human Rights, Democracy and the Rule of Law'. Report Prepared in the Context of the Council of Europe's Ad Hoc Committee on AI (CAHAI). Strasbourg: Council of Europe, 24 June 2020. https://www.coe.int/en/web/artificial-intelligence/cahai.

Murray, Hugh. Review of *Review of Modernity and the Holocaust*, by Zygmunt Bauman. *German Politics & Society*, no. 22 (1991): 82–86.

Norris, Pippa. 'The Populist Challenge to Liberal Democracies'. In *The Oxford Handbook of Political Representation in Liberal Democracies*, edited by Robert Rohrschneider and Jacques Thomassen, 543–62. Oxford University Press, 2020. https://doi.org/10.1093/oxfordhb/9780198825081.013.28.

Ntoutsi, Eirini, Pavlos Fafalios, Ujwal Gadiraju, Vasileios Iosifidis, Wolfgang Nejdl, Maria-Esther Vidal, Salvatore Ruggieri, et al. 'Bias in Data-Driven Artificial Intelligence Systems—An Introductory Survey'. *WIREs Data Mining and Knowledge Discovery* 10, no. 3 (2020): e1356. https://doi.org/10.1002/widm.1356.

Ochigame, Rodrigo. 'The Invention of "Ethical AI": How Big Tech Manipulates Academia to Avoid Regulation'. *The Intercept* (blog), 20 December 2019. https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/.

OECD. 'Recommendation of the Council on Artificial Intelligence', May 2019. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

O'Neil, Cathy. *Weapons of Math Destruction*. Penguin Books Ltd, 2017.

Pasquale, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press, 2015. http://www.jstor.org/stable/j.ctt13x0hch.

Pollock, Benjamin. ‘Franz Rosenzweig’. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2019. Metaphysics Research Lab, Stanford University, 2019. https://plato.stanford.edu/archives/spr2019/entries/rosenzweig/.

Rességuier, Anaïs, and Rowena Rodrigues. ‘AI Ethics Should Not Remain Toothless! A Call to Bring Back the Teeth of Ethics’. *Big Data & Society* 7, no. 2 (1 July 2020): 2053951720942541. https://doi.org/10.1177/2053951720942541.

Robinson, George. *Essential Judaism: A Complete Guide to Beliefs, Customs & Rituals*. New York: Atria Books, 2016.

Rosenzweig, Franz. *Der Stern Der Erlösung*. Translated by Alexandre Derczanski and Jean-Louis Schlegel. Paris: Editions du Seuil (2003), 1976.

Rostami, Hamidey, Jean-Yves Dantan, and Lazhar Homri. ‘Review of Data Mining Applications for Quality Assessment in Manufacturing Industry: Support Vector Machines’. *International Journal of Metrology and Quality Engineering* 6, no. 4 (2015). http://dx.doi.org.kuleuven.e-bronnen.be/10.1051/ijmqe/2015023.

Russell, Stuart Jonathan, and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Fourth edition. Pearson Series in Artificial Intelligence. Hoboken: Pearson, 2021.

Sacks, Samm, and Justin Sherman. ‘Calling Data “the New Oil” Could Hurt Efforts to Protect Privacy’. *Slate Magazine* (blog), 13 June 2019. https://slate.com/technology/2019/06/data-not-new-oil-kai-fu-lee-china-artificial-intelligence.html.

Salles, Arleen, Kathinka Evers, and Michele Farisco. ‘Anthropomorphism in AI’. *AJOB Neuroscience* 11, no. 2 (2 April 2020): 88–95. https://doi.org/10.1080/21507740.2020.1740350.

Scholem, Gershom. *On Jews and Judaism in Crisis: Selected Essays*. Edited by Werner J. Dannhauser. New York: Schocken Books, 1976.

Sharpe, Matthew. ‘When the Logics of the World Collapse - Zizek with and against Arendt on “Totalitarianism”’. *Subjectivity* 3, no. 1 (April 2010): 53–75.

Shi, Chenyu, Joost M. Meijer, George Azzopardi, Gilles F. H. Diercks, Jiapan Guo, and Nicolai Petkov. ‘Use of Convolutional Neural Networks for the Detection of U-Serrated Patterns in Direct Immunofluorescence Images to Facilitate the Diagnosis of Epidermolysis Bullosa Acquisita’. *The American Journal of Pathology*, 28 June 2021. https://doi.org/10.1016/j.ajpath.2021.05.024.

Simon, Thomas W. *Democracy and Social Injustice: Law, Politics, and Philosophy*. Rowman & Littlefield, 1995.

Sloot, Bart van der, and Sascha van Schendel. ‘Procedural Law for the Data-Driven Society’. *Information & Communications Technology Law* (20 January 2021): 1–29. https://doi.org/10.1080/13600834.2021.1876331.

Smuha, Nathalie A. ‘Artificiële intelligentie bij de overheid. Opportuniteiten en uitdagingen vanuit ethisch-juridisch perspectief’. *Vlaams Tijdschrift voor Overheidsmanagement (VTOM)*, no. 4 (2019).

———. ‘Beyond a Human Rights-Based Approach to AI Governance: Promise, Pitfalls, Plea’. *Philosophy & Technology*, 24 May 2020. https://doi.org/10.1007/s13347-020-00403-w.

———. ‘Beyond the Individual: Governing AI’s Societal Harm’. *Internet Policy Review*, 10(3), 2021.

———. ‘From a “Race to AI” to a “Race to AI Regulation”: Regulatory Competition for Artificial Intelligence’. *Law, Innovation and Technology* 13, no. 1 (2 January 2021): 57–84. https://doi.org/10.1080/17579961.2021.1898300.

———. ‘Laten We Intelligenter Zijn Wanneer We Het over Artificiële Intelligentie Hebben’. Knack Data News, 11 March 2020. https://datanews.knack.be/ict/nieuws/laten-we-intelligenter-zijn-wanneer-we-het-over-artificiele-intelligentie-hebben/article-opinion-1574905.html.

———. ‘The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence’. *Computer Law Review International* 20, no. 4 (2019): 97–106.

———. ‘Trustworthy Artificial Intelligence in Education: Pitfalls and Pathways’. Social Science Research Network, 2020. https://doi.org/10.2139/ssrn.3742421.

Smuha, Nathalie A., Emma Ahmed-Rengers, Adam Harkens, Wenlong Li, James MacLaren, Riccardo Piselli, and Karen Yeung. ‘How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission’s Proposal for an Artificial Intelligence Act’. Social Science Research Network, 5 August 2021. https://papers.ssrn.com/abstract=3899991.

Somers, James. ‘The Pastry A.I. That Learned to Fight Cancer’. *The New Yorker*. 18 March 2021. https://www.newyorker.com/tech/annals-of-technology/the-pastry-ai-that-learned-to-fight-cancer.

Spaid, Sue. ‘Surfing the Public Square: On Worldlessness, Social Media, and the Dissolution of the Polis’. *Open Philosophy* 2, no. 1 (31 December 2019): 668–78.

Stark, Luke. ‘Algorithmic Psychometrics and the Scalable Subject’. *Social Studies of Science* 48, no. 2 (April 2018): 204–31. https://doi.org/10.1177/0306312718772094.

Strubell, Emma, Ananya Ganesh, and Andrew McCallum. ‘Energy and Policy Considerations for Deep Learning in NLP’. *ArXiv:1906.02243 [Cs]*, 5 June 2019. http://arxiv.org/abs/1906.02243.

Suresh, Harini. ‘The Problem with “Biased Data”’. Medium, 26 April 2019. https://medium.com/@harinisuresh/the-problem-with-biased-data-5700005e514c.

Svoboda, Elizabeth. ‘Artificial Intelligence Is Improving the Detection of Lung Cancer’. *Nature* 587, no. 7834 (18 November 2020): S20–22. https://doi.org/10.1038/d41586-020-03157-9.

Taylor, Linnet. ‘Public Actors Without Public Values: Legitimacy, Domination and the Regulation of the Technology Sector’. *Philosophy & Technology*, 20 January 2021. https://doi.org/10.1007/s13347-020-00441-4.

Taylor, Linnet, Luciano Floridi, and Bart van der Sloot, eds. *Group Privacy: New Challenges of Data Technologies*. Cham: Springer International Publishing, 2017. https://doi.org/10.1007/978-3-319-46608-8.

Theodorou, Andreas, and Virginia Dignum. ‘Towards Ethical and Socio-Legal Governance in AI’. *Nature Machine Intelligence* 2, no. 1 (January 2020): 10–12. https://doi.org/10.1038/s42256-019-0136-y.

Thomas, Rachel, and David Uminsky. ‘The Problem with Metrics Is a Fundamental Problem for AI’. *Ethics of Data Science Conference 2020*, 19 February 2020. http://arxiv.org/abs/2002.08512.

Thompson, Dennis F. ‘Designing Responsibility: The Problem of Many Hands in Complex Organizations’. In *Designing in Ethics*, edited by Jeroen van den Hoven, Seumas Miller, and Thomas Pogge, 1st ed., 32–56. Cambridge University Press, 2017. https://doi.org/10.1017/9780511844317.003.

Tomašev, Nenad, Julien Cornebise, Frank Hutter, Shakir Mohamed, Angela Picciariello, Bec Connelly, Danielle C. M. Belgrave, et al. ‘AI for Social Good: Unlocking the

Opportunity for Positive Impact'. *Nature Communications* 11, no. 1 (18 May 2020): 2468. https://doi.org/10.1038/s41467-020-15871-z.

Topolski, Anya. *Arendt, Levinas and a Politics of Relationality*. Reframing the Boundaries: Thinking the Political. Lanham: Rowman and Littlefield, 2015.

Turing, Alan. 'Computing Machinery and Intelligence'. *Mind* 59, no. 236 (October 1950): 433–60.

UK Department for Education. 'Realising the Potential of Technology in Education: A Strategy for Education Providers and the Technology Industry', 2019. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/791931/DfE-Education_Technology_Strategy.pdf.

UNESCO Ad Hoc Expert Group (AHEG) for the Preparation of a Draft text of a Recommendation the Ethics of Artificial Intelligence. 'First Version of a Draft Text of a Recommendation on the Ethics of Artificial Intelligence', 2020. https://unesdoc.unesco.org/ark:/48223/pf0000373434.

Varona, Daniel, Yadira Lizama-Mue, and Juan Luis Suárez. 'Machine Learning's Limitations in Avoiding Automation of Bias'. *AI & Society* 36, no. 1 (March 2021): 197–203. https://doi.org/10.1007/s00146-020-00996-y.

Veale, Michael. 'A Critical Take on the Policy Recommendations of the EU High-Level Expert Group on Artificial Intelligence'. *European Journal of Risk Regulation*, 23 January 2020, 1–10. https://doi.org/10.1017/err.2019.65.

Vergauwen, Roger. 'Will Science and Consciousness Ever Meat? Complexity, Symmetry and Qualia'. *Symmetry* 2, no. 3 (September 2010): 1250–69. https://doi.org/10.3390/sym2031250.

Verovšek, Peter J. 'Integration after Totalitarianism: Arendt and Habermas on the Postwar Imperatives of Memory'. *Journal of International Political Theory* 16, no. 1 (1 February 2020): 2–24. https://doi.org/10.1177/1755088218796535.

Viljoen, Salomé. 'Democratic Data: A Relational Theory for Data Governance'. *Available at SSRN: Https://Ssrn.Com/Abstract=3727562*, November 2020.

Vinuesa, Ricardo, Hossein Azizpour, Iolanda Leite, Madeline Balaam, Virginia Dignum, Sami Domisch, Anna Felländer, Simone Daniela Langhans, Max Tegmark, and Francesco Fuso Nerini. 'The Role of Artificial Intelligence in Achieving the Sustainable Development Goals'. *Nature Communications* 11, no. 1 (13 January 2020): 233. https://doi.org/10.1038/s41467-019-14108-y.

Wagner, Ben. 'Ethics As An Escape From Regulation. From "Ethics-Washing" To Ethics-Shopping?' In *Being Profiled*, edited by Emre Bayamlioglu, Irina Baraliuc, Liisa Albertha Wilhelmina Janssens, and Mireille Hildebrandt, 84–89. Amsterdam: Amsterdam University Press, 2019. https://doi.org/10.1515/9789048550180-016.

Walker, Kent, and Fang Wan. 'The Harm of Symbolic Actions and Green-Washing: Corporate Actions and Communications on Environmental Performance and Their Financial Implications'. *Journal of Business Ethics* 109, no. 2 (2012): 227–42.

Wang, Guan, Yu Sun, and Jianxin Wang. 'Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning'. *Computational Intelligence and Neuroscience* 2017 (5 July 2017): e2917536. https://doi.org/10.1155/2017/2917536.

Warren, Samuel D., and Louis D. Brandeis. 'The Right to Privacy'. *Harvard Law Review* 4, no. 5 (1890): 193–220.

Werpachowska, Agnieszka. '"Computer Says No": Was Your Mortgage Application Rejected Unfairly?' *Wilmott* 2020, no. 108 (2020): 54–61. https://doi.org/10.1002/wilm.10858.

Winner, Langdon. ‘Do Artifacts Have Politics?’ *Daedalus* 109, no. 1 (1980): 17.

Wohl, Benjamin S. ‘Revealing the “Face” of the Robot: Introducing the Ethics of Levinas to the Field of Robo-Ethics’. In *Mobile Service Robotics*, 704–14. Poznan, Poland: World Scientific, 2014.

Yanklowitz, Shmuly. *Pirkei Avot: A Social Justice Commentary*. New York: CCAR Press, 2018.

Yearsley, Liesl. ‘We Need to Talk about the Power of AI to Manipulate Us’. MIT Technology Review. Accessed 12 November 2019. https://www.technologyreview.com/s/608036/we-need-to-talk-about-the-power-of-ai-to-manipulate-humans/.

Yeung, Karen. ‘Five Fears about Mass Predictive Personalization in an Age of Surveillance Capitalism’. *International Data Privacy Law* 8, no. 3 (1 August 2018): 258–69. https://doi.org/10.1093/idpl/ipy020.

———. ‘Responsibility and AI - A Study of the Implications of Advanced Digital Technologies (Including AI Systems) for the Concept of Responsibility within a Human Rights Framework’. Council of Europe, DGI(2019)05, September 2019.

———. ‘Why Worry about Decision-Making by Machine?’ In *Algorithmic Regulation*, edited by Karen Yeung and Martin Lodge, 21–48. Oxford University Press, 2019. https://doi.org/10.1093/oso/9780198838494.003.0002.

Yeung, Karen, Andrew Howes, and Ganna Pogrebna. ‘AI Governance by Human Rights–Centered Design, Deliberation, and Oversight: An End to Ethics Washing’. In *The Oxford Handbook of Ethics of AI*, edited by Markus D. Dubber, Frank Pasquale, and Sunit Das, 75–106. Oxford University Press, 2020. https://doi.org/10.1093/oxfordhb/9780190067397.013.5.

Zambak, Aziz, and Roger Vergauwen. ‘Artificial Intelligence and Agentive Cognition: A Logico-Linguistic’. *Logique et Analyse* 52, no. 205 (2009): 57–96.

Zheng, Huadi, Haibo Hu, and Ziyang Han. ‘Preserving User Privacy for Machine Learning: Local Differential Privacy or Federated Machine Learning?’ *IEEE Intelligent Systems* 35, no. 4 (2020): 5–14. https://doi.org/10.1109/MIS.2020.3010335.

Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. 1st ed. New York: PublicAffairs, 2019.

Zuiderveen Borgesius, Frederik J. ‘Discrimination, Artificial Intelligence, and Algorithmic Decision-Making’. Strasbourg: Council of Europe - Directorate General of Democracy, 2018.

———. ‘Strengthening Legal Protection against Discrimination by Algorithms and Artificial Intelligence’. *The International Journal of Human Rights*, 25 March 2020, 1–22. https://doi.org/10.1080/13642987.2020.1743976.

Zuiderveen Borgesius, Frederik J., Judith Möller, Sanne Kruikemeier, Ronan Ó Fathaigh, Kristina Irion, Tom Dobber, Balazs Bodo, and Claes De Vreese. ‘Online Political Microtargeting: Promises and Threats for Democracy’. *Utrecht Law Review* 14, no. 1 (9 February 2018): 82. https://doi.org/10.18352/ulr.420.