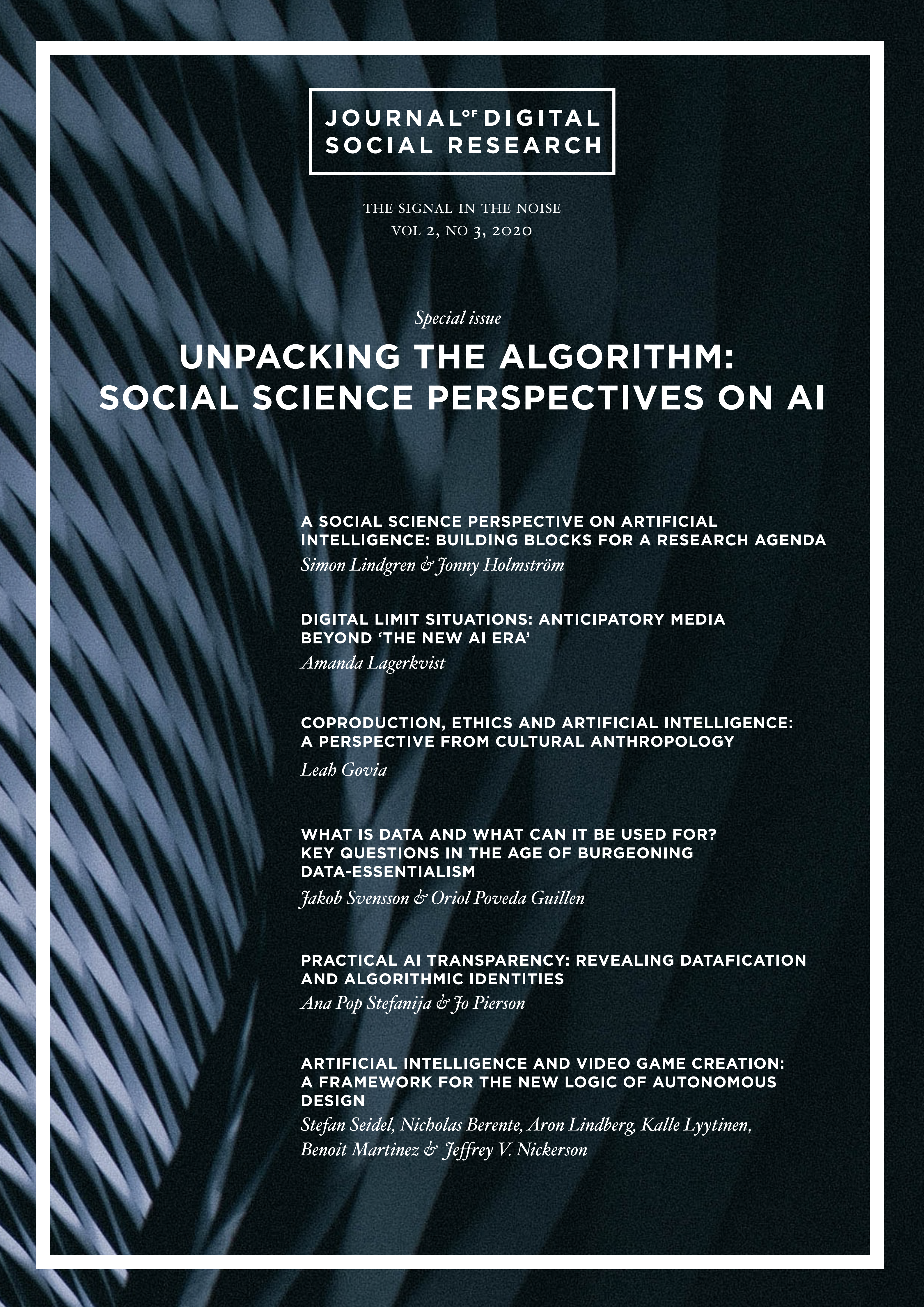


JOURNAL OF DIGITAL SOCIAL RESEARCH

THE SIGNAL IN THE NOISE
VOL 2, NO 3, 2020

Special issue

UNPACKING THE ALGORITHM: SOCIAL SCIENCE PERSPECTIVES ON AI

JOURNAL OF DIGITAL SOCIAL RESEARCH

VOL. 2 : NO. 3 : 2020

EDITORIAL BOARD

JOURNAL METADATA

ISSN
2003–1998

DOI
http://doi.org/10.33621/jdsr

WEB
www.jdsr.io

ABOUT JDSR

JDSR is a cross-disciplinary, online, open-access journal, focussing on the interaction between digital technologies and society. JDSR is published by DIGSUM, the Centre for Digital Social Research at Umeå University, Sweden.

CONTACT US

E-MAIL
editor@jdsr.se

INTERNATIONAL EDITORIAL BOARD

LICENCE & COPYRIGHT

Cover photo by Victor Garcia, Unsplash.com

JOURNAL OF DIGITAL SOCIAL RESEARCH

VOL. 2 : NO. 3 : 2020

Special issue

UNPACKING THE ALGORITHM: SOCIAL SCIENCE PERSPECTIVES ON AI

JOURNAL OF DIGITAL SOCIAL RESEARCH

WWW.JDSR.IO ISSN: 2003-1998

VOL. 2, NO. 3, 2020, 1–15

A SOCIAL SCIENCE PERSPECTIVE ON ARTIFICIAL INTELLIGENCE: BUILDING BLOCKS FOR A RESEARCH AGENDA

Simon Lindgren and Jonny Holmström*

ABSTRACT

In this article, we discuss and outline a research agenda for social science research on artificial intelligence. We present four overlapping building blocks that we see as keys for developing a perspective on AI able to unpack the rich complexities of sociotechnical settings. First, the interaction between humans and machines must be studied in its broader societal context. Second, technological and human actors must be seen as social actors on equal terms. Third, we must consider the broader discursive settings in which AI is socially constructed as a phenomenon with related hopes and fears. Fourth, we argue that constant and critical reflection is needed over how AI, algorithms and datafication affect social science research objects and methods. This article serves as the introduction to this JDSR special issue about social science perspectives on AI.

Keywords: social science; artificial intelligence; sociotechnical perspectives; social constructionism

* Umeå University, Sweden.

1 TOWARDS A TRULY SOCIAL SCIENCE OF AI

People have been grappling with the social consequences of technology for centuries. Take, for example, Langdon Winner's (1980) example of how New York City's overpasses were built, in the early-to-mid 20th century, in ways that discouraged the presence of buses on the parkways. This was analysed, later on, as a result of master builder Robert Moses' racial prejudice and social-class bias. While the design of the overpasses allowed car-owning whites of the upper and middle classes to use them for recreation and commuting, low-income groups and racial minorities – who largely relied on public transport – were effectively denied access (Woolgar, & Cooper, 1999). Such examples clearly illustrate that technologies are political. They embody power and social relations. Historian of technology Melvin Kranzberg (1986, p. 545–546) has argued that:

> Technology is neither good nor bad; nor is it neutral. […] Technology's interaction with the social ecology is such that technical developments frequently have environmental, social, and human consequences that go far beyond the immediate purposes of the technical devices and practices themselves, and the same technology can have quite different results when introduced into different contexts or under different circumstances.

As technology is political, and because it is preceded, succeeded, and surrounded by the social, the comprehensive study of any technologies, including artificial intelligence (AI), demands a social science perspective.

In a recent paper on the emerging scholarly field of *machine behaviour*, Rahwan et al. (2019) point out the fact that AI is still predominantly studied by the same scientists who are engaged in creating the AI agents themselves. This leads to a strong focus on research that in various ways is designed to ensure that AI fulfils intended functions. AI is seen as having to be adequate, efficient, responsible, and so on. And even though it could be argued that social scientists, and also humanities scholars, are taking part in AI research to a growing degree (Araujo et al., 2020; Dung et al., 2020; Gupta & Tu, 2020; Miller et al., 2017), the research agenda is still largely set through posing questions based in the AI technologies per se, rather than in their social and cultural contexts.

In many cases, an interdisciplinary approach to the study of AI is advisable. Social scientists can clearly learn a lot about technological aspects of AI from those that work with developing AI systems, agents, and algorithms, and such understanding is key to carrying out well-informed research on the societal dimensions of these (Reutter, 2018; Richardson, 2015). Conversely, computer scientists and AI developers can get valuable knowledge through carrying out user-studies and evaluations of implemented systems by collaborating with social scientists (Guzman, 2017; Irving & Askell, 2019), as this can counteract "AI's social sciences deficit" (Sloane & Moss, 2019).

Even if code is social, and the social is code, the purely technological sciences must in some instances detach themselves from social and cultural considerations to work simply on the technological side of AI – on the silicon and digits. And just as well, the social sciences must sometimes disconnect from technological considerations to focus on purely socio-cultural dimensions of AI. In spite of the many advantages of interdisciplinary research, there is often a translation problem between social and technological research, also potentially involving a mismatch between different overarching objectives for the research as such. As some social science research has shown, "AI doesn't make everybody's life easier or safer" (Sloane & Moss, 2019, p. 330). It can also exacerbate inequality, lead to discrimination, and inflict harm based on race, gender and class (Eubanks, 2017; Noble, 2018; O'Neil, 2016).

While research on how AI systems may reproduce, or sometimes even worsen, prevailing patterns of oppression, can function as direct input into work with enhancing AI technologies themselves, this must not always be the case. Another, even more important, role for social science research is to do what it does best, that is systematically analyse society, scrutinize historical continuities and discontinuities, and to produce knowledge about the political, economic and social structures and conditions under which we live. This includes focusing on issues of power and oppression, on social differences, on identities, on language and ideologies, and on hindrances or possibilities for action for given individuals and collectives. Such knowledge has a value in itself, and as indirect input into a broad range of other scholarly fields.

In light of this, we argue in favour of proliferating, alongside relevant efforts to evaluate the social consequences of particular AI technologies, a truly social science of AI as a political and socio-historical phenomenon. This entails drawing on well-established literatures in the social sciences which relates to (1) *Humans and machines in context*, (2) *AI agents as social actors*, (3) *AI as social construction*, and (4) *AI, datafication and research methods.*

1.1 Humans and machines in context

Practices and concepts for understanding the role of code and software in human-computer interaction (HCI) have been developed in literature from computer and information sciences since the coming of personal computing in the 1980s (Dix 2004; Preece 1994). HCI scholars like Suchman (1987; 2009) and Nardi (1995) have emphasized the importance of taking contextual and socio-cultural dimensions into account and have argued for a view where humans and machines constantly construct and reconstruct the social world through dynamic interactions. These perspectives have been influential in areas where the aim has been to improve the usability of computers, and designing systems that make human-computer interaction flow as smoothly as possible

Retaining the key idea in HCI, that communication between humans and machines is a socio-cultural rather than a technological process (cf. Carey, 2009), we suggest that social science research on AI must move far past issues of mere usability, fairness, and responsibility, towards a research framework that allows for posing more far-reaching and deeper-cutting questions. A promising path would be to position social research on AI closer to the area of human-machine communication (HMC) as outlined by scholars such as Guzman and Lewis (2019; Guzman 2018). This pushes in the direction of conceiving of AI agents not as mere AI technologies, but as communicative agents that engage in ongoing and adaptive acts of communication in people's everyday social spaces. This view challenges many concepts that tend to be taken for granted in social science research, such as the question of what constitutes an actor (cf. Latour, 2005). But as is ever more evident from the emerging and proliferated presence of AI in public and civic life, interaction and communication can no longer be seen as a human-only process. Instead, we must accommodate the study of the interplay between people, and between people and AI, within one and the same theoretical framework. How can we best account for social structures that also include social machines?

While one strategy is to simply understand AI agents as technologically "automated social actors" (Abokhodair et al., 2015), AI is created by humans and thus encoded with human intentions (Siponen, 2004). This means that they embody social values, which makes them human-dependent rather than completely autonomous (Keller & Klinger, 2019). Furthermore, the behaviour of AI systems can be affected by input from the humans with whom they interact. Generally, as argued by Carey (1990, p. 247), technologies always function as "concrete embodiments of human purposes, social relations, and forms of organization". As AI is always somehow imbued with social intentionality, is must also be seen as a site of power (Chun, 2011; Holmström & Robey, 2020). The path for social theory past this increasingly altered border between human and machine goes through assuming a hybrid, or 'cyborg', perspective.

1.2 AI agents as social actors

A truly social science of AI needs to approach the human/AI relationship as complex and multidimensional (Gehl & Bakardjieva, 2017). This means expecting a symbiotic interconnection between technological and human elements (Neff & Nagy, 2016). Drawing on the perspective of Carey (2009), AI must be seen as being simultaneously constituted and expressed in an ongoing relationship with a surrounding social world (Carey, 1990, p. 247). We believe that in order to be able to demystify AI as an analytical category (cf. Barocas et al., 2013), we must study its agents alongside other agents in their social and communicative context.

This perspective also aligns with the constructionist view on technology and society which is widely advocated in the field of Science and Technology Studies (STS). In this field, sociologists such as Latour, Callon, and Law have contributed to formulating so-called Actor-Network theory (ANT), that allows for networks of social action where the agency of human and non-human agents is seen as equal (Bijker & Law, 1992; Callon, 1986; Latour & Callon, 1991). This is an analytical approach that wants to move beyond the anthropological, human-centred, bias of traditional sociology and instead focus on the entangled and symbiotic nature of relationships between humans and technologically social actors such as for example software, algorithms, and intelligent agents (Faraj et al. 2018; Woolley, 2018).

Social relations can emerge between all different kinds of entities — which is what happens when the actions of one entity (e.g. a bot or a human) has an effect on the actions of another (e.g. a human or a bot). Because of this, a social science approach to AI has much to gain by drawing on theories such as ANT insofar that it provides conceptual tools for exploring the complex role of intelligent agents in online and offline socio-technical systems. The key concept within ANT, which fits the most succinctly here comes from Latour's discussions of how technological artifacts can both replace human actions and shape further human actions. His notion of "delegation" refers to processes where human agents, such as for example engineers, design technological systems to which they *delegate* tasks to be carried out on human behalf. Latour points out how us humans "have been able to delegate to nonhumans not only force as we have known it for centuries but also values, duties, and ethics" (Latour, 1992, p. 232). We can conceive of AI as technologies to which human subjects delegate agency and abilities. In turn, these "non-humans intervene actively to push action in unexpected directions" (Callon & Law, 1997, p. 178).

1.3 AI as social construction

In approaching AI as an object of social scientific study, it is useful to draw on understandings that have been developed within the theoretical tradition referred to as the *social shaping of technology* (Pinch & Bijker, 1984; Williams & Edge, 1996). Aligning with ideas developed in this area, we see the social scientific study of AI as by necessity having a strong discursive component, focusing on how social talk and action around it is structured.

The design and implementation of technologies is always socially and historically dependent, and technologies are used and developed in processes that are based on a variety of social considerations (MacKenzie & Wajcman, 1985).

The socio-technical phenomenon of 'AI' comes into being through a co-construction process where various *interpretive frames* are negotiated and established. According to scholars in this field, technologies are surrounded by "socially shared structures of meaning" (Latzko-Toth, 2014, p. 50), that reflect and orient how various groups of actors relate to a given technological artifact and how they make sense of it. Bijker (1987) argues that such modes of speaking and acting in relation to technological artifacts constitutes an interpretive "frame", that provides a "grammar" for how meaning is attributed to the artifact in question. Such frames include "assumptions, knowledge, and expectations, expressed symbolically through language, visual images, metaphors, and stories" (Orlikowski & Gash, 1994, p. 178). These will have powerful effects, as the knowledge, assumptions, and expectations that people have about the meaning, purpose, and importance of technology will influence their societal uses and hence their impact. Another way of putting this is that the interpretive frames will affect how the technology in question becomes *socialized* — how it becomes a social object and how it acquires social significance (Jouet, 2000; Scott & Orlikowski, 2014), as the result of a process where its relevance, meaning and compatibility with societal norms and values are negotiated and debated (Latzko-Toth, 2014; Mallein & Toussaint, 1994).

1.4 AI, datafication and research methods

More broadly speaking, we conceive of AI agents and technologies as being part of — and an expression of — the social and communicative hybridity that is characteristic of 21st century society (Chadwick, 2013; Lindgren, 2014). Carrying out truly social science on AI, therefore, must also take into account and investigate the role and impact of networks, software, and algorithms on the social, cultural, and political. AI, its developers, and subjects/users, analysed in context can be considered to be "hybrid techno-social formations" (Woolley, 2018, p. 134). A central aspect of these formations is their *datafication*, a process that not only affects society at large and comprehensively, and which supplies some present-day AI with crucial raw material, but that also has an impact on our choice of research methods and analytical strategies when studying AI as social scientists. Datafication is the process which has led to the situation where we now live in "a culture that is shaped and populated with numbers, where trust and interest in anything that cannot be quantified diminishes" (Beer, 2016, p. 149). Furthermore, in the age of big data, there is an obsession with causation. As boyd and Crawford (2012, p. 665) argue, the mirage and mythology of big data demand that a number of critical questions are raised with regard to "what all this data means, who gets access to what data, how data analysis is employed, and to what ends". There is a risk that the lure of big data will sideline other forms of analysis, and that other alternative methods with which to analyse the beliefs, choices, expressions, and strategies of people are pushed aside by the sheer volume of numbers.

We believe that a truly social science of AI, must rely on a custom and open-minded combinations of methodological and theoretical approaches (cf. Lindgren, 2020). This means sometimes embracing both the massive flows of data, as well as computational analytical approaches, and sometimes stepping out of the data flows, observing them through the lens of tried and tested social and cultural theories about technology and social change, or other critical perspectives. This also entails approaching the digital object of study through forms of hermeneutic, ethnographic, and seemingly 'analogue' methods. These convictions position our suggested perspective in a sympathetic position in relation to the area of *software studies* (Kitchin & Dodge, 2011; Manovich, 2013), that is focused on studying various social expressions of computer code as politically imbued and analysing how algorithmic agency is entangled with social practice (Gillespie, 2014, p. 168). As Lindgren (2020, p. 12) writes:

> Being data-driven is not a bad thing, but there must always be a balance between data and theory – between information and its interpretation. This is where sociology and social theory come into the picture, as they offer a wide range of conceptual frameworks, theories, that can aid in the analysis and understanding of the large amounts and many forms of social data that are proliferated in today's world.

AI, in its full sense, is only partly a technological phenomenon. It is also a cultural and socio-political phenomenon, imbued with certain assumptions, hopes, beliefs, and ideologies. The consequences of AI span a range of areas, including challenges as well as opportunities relating to power, oppression, health, work, economy, sustainability, learning, inclusion, diversity, and justice. Prominently, AI and automated agents also play into processes of democracy, governance, and social trust. This development, where the emergence and proliferation of AI agents based on algorithms are key, most definitely demands to be scrutinised from a social science perspective. We need more knowledge about what the pervasive use of these human-software hybrids, and the black-boxed and sometimes discriminatory algorithms behind them, mean for future societies. Critical social science research must run alongside and monitor the development by which AI agents will unavoidably become increasingly interwoven in our society, in areas ranging from online dating and credit scoring, through parenting and education, to social welfare control, policing and warfare.

2 ARTIFICIAL INTELLIGENCE: BUILDING BLOCKS IN THE SOCIAL SCIENCE APPROACH

Researchers have begun to address the real-life quandaries that AI introduces (e.g. Boden, 2016; Bostrom, 2016). But while we are thrilled to see how some AI researchers are increasingly addressing the legal, political, economic and societal aspects of AI, we are surprised over the ways in which many technology-focused AI researchers tend to ignore decades of social science technology research. We are equally worried over how social scientists have been slow starters in researching AI. This has meant that scholars that lack the appropriate expertise have begun to take on social questions on their own, without any solid foundation in social science. At the same time, scholars from the social sciences, physical sciences, and humanities seem to be losing touch with the rapid advances in AI (Frank et al., 2019). We believe that the contributions to this special issue of JDSR are illustrative examples of how AI can be approached in ways that include a strong social science element.

As a first building block, we wrote above about the importance of looking at *humans and machines in context*. The social scientific study of AI is interested in how humans and machines interact to construct their social world. Machines are technological, humans are social, but in context they are socio-cultural phenomena. This perspective must go both ways, recognising the agency, as well as the structurally defined ('programmed', as it were) character of humans as well as machines. In this special issue such a contextual perspective comes to the fore in Govia's (2020) contribution. This study targets assumptions of technological determinism and shifts focus to everyday interaction with AI systems and processes. Fruitfully drawing on an STS perspective, Govia contributes to a situated understanding of AI. Similarly, in another contribution, Seidel et al. (2020) write about how AI use in video game creation can be analysed. They apply a contextualised perspective where the autonomous design tools are seen as participating agents in the design process, and also draw on control theory to analyse the relationship between context, humans, and technology.

Our second building block was about approaching and conceptualising *AI agents as social actors*. In doing so we also pointed to the usefulness of applying an STS perspective, such as Govia's, according to which the agency of human and non-human (e.g. technological) actants are seen as equal. Like Govia's rendition of anthropology, this view wants to move away from human-centred social science towards more entangled ways of seeing humans and technologically social actors (software, algorithms, intelligent agents). In Svensson and Poveda Guillen's (2020) contribution to this issue, the authors align with a view of data and algorithms as dynamic actants, rather than as objective and firmly-set entities. They develop a compelling critique of data-essentialism and contend that seeing both the data/algorithms and their human subjects as dynamic and historically shaped, can counteract the rise of a new form of positivism. Connecting also to our previous point about the importance of context, the authors write that:

> [A]cknowledging the importance of data, conceiving of data as contextual and situated traces we leave behind in an increasingly computer saturated world is substantially different from reducing our existence and bodies to data (Svenson & Poveda Guillen, 2020, p. 78).

Both of the above points, focusing on the embedded and interactive character of AI as phenomenon, in turn relates to our third building block about *AI as a social construction*. In our discussion of this point, we especially emphasised that AI has a strong discursive component, meaning that it is, like so many other terms, part of a political language. It gets shaped, defined, and acquires its social significance through how it is framed and understood, and through which hopes or fears are symbolically tied to it. Svensson and Poveda Guillen's paper in this special issue is also of strong relevance to this, as it critiques how data tends to be seen as objective, and suggests alternative views. Digging deeper into this territory, Lagerkvist's (2020) contribution draws on the existential philosophy of Karl Jaspers to discuss how AI is not merely a medium, but also a message. Addressing similar discursive issues as those mentioned above, Lagerkvist problematises how the self-presentation of AI mythologically constructs its futures as inevitable. This is not 'simply' about talk and discourse, as the current moment, Lagerkvist argues, constitutes a "digital limit situation" with high political and ethical stakes. The stakes are also existential, as the ways in which AI futures are imagined symbolically close down other potential futures. Lagerkvist writes:

> Presenting themselves as the only set of solutions to problems that face us on the fringes of our late modern societal order of disintegration – while operating through forecasting, prediction and precision – [AI imaginaries] thus effectively close the very horizon of the future at the same time (Lagerkvist, 2020, p. 35).

It is through such theoretical insights, and through empirical research that draws upon them, that social and cultural perspectives on AI can make important contributions. Sometimes social scientists can help evaluate whether this or that AI system is more or less user-friendly, more or less democratic, or more or less accurate, or more or less ethical. This is equal to doing social science research within the paradigm of what Lagerkvist calls the prevailing 'AI imaginaries', and often drawing on what Svensson and Poveda Guillen label as 'data-essentialism'. The truly social science of AI, especially a critical one, lies beyond such confines and enables posing questions not only from inside the technological paradigm, but from the outside.

Doing such work entails a range of methodological challenges, the depth and scope of which exceeds what we can address within this special issue alone. However, Pop Stefanija and Pierson's (2020) contribution, to this special issue addresses some of the challenges with researching algorithms from the outside in the face of their inherent opacity and black-boxedness. The issues that they discuss relate to our fourth building block presented earlier, namely that of dealing with *AI and datafication in relation to research methods*. Stefanija and Pierson discuss a number of limitations with API-based research, and how constant changes in platforms' politics of visibility constitutes data access gaps. The authors' work is an enlightening example of how the present data landscape demands continuous adaptation and smart combinations of both new and existing methods. Pop Stefanija and Pierson advocate an approach, using non-traditional research tools, in an endeavour to letting and making 'the platforms speak'.

3 ARTIFICIAL INTELLIGENCE AND THE SOCIAL SCIENCES: TOWARDS A RESEARCH AGENDA

AI is a rapidly emerging phenomenon of societal significance. As such, the ethical and social implications of AI have become topics of compelling interest to academia, industry, and the public. We however find that the dominant framings of AI are still limited since they tend to approach AI in a narrow and deterministic way, essentially understanding AI as a shaper of society. The examples of social science perspectives on AI in this special issue together demonstrate a richer and more multifaceted view, in which AI is indeed seen to shape society, but not necessarily in the ways envisioned by its creators, and where society's shaping of AI is also highlighted.

A social science research agenda on AI should be informed by such a mutual shaping approach guiding our inquiry into the dynamic processes of AI design and use, suggesting that society and AI are not mutually exclusive but, instead, influence and shape each other. As a whole, the papers in this special issue demonstrate, in different ways, what is to be gained from a applying a mutual shaping approach, and to focus on analysing how social and cultural factors influence the ways in which technologies are designed, used, and evaluated, as well as how technologies affect the construction of society.

AI is still a poorly understood societal phenomenon today, since social scientists have been slow out of the gates. By building on the rich resources we find in social science theory and method, we can articulate a truly social science approach to AI. Again, as building blocks in such a social science approach, we suggest considering: (1) Humans and machines in context, (2) AI agents as social actors, (3) AI as social construction, and (4) AI's relationship to datafication and research methods. By drawing on these building blocks, social science scholars can fruitfully explore the complexities involved in human-machine configurations to contribute to the emerging scholarly AI discourse.

REFERENCES

Abokhodair, N., Yoo, D. and McDonald, D. W. (2015) 'Dissecting a social botnet: Growth, content and influence in Twitter', in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, pp. 839–851.

Araujo, T. *et al.* (2020) 'In AI we trust? Perceptions about automated decision-making by artificial intelligence', *AI & SOCIETY*, 35(3), pp. 611–623. doi: 10.1007/s00146-019-00931-w.

Barocas, S., Hood, S. and Ziewitz, M. (2013) *Governing Algorithms: A Provocation Piece*. SSRN scholarly paper ID 2245322. Rochester, NY: Social Science Research Network. Available at: https://papers.ssrn.com/abstract=2245322 (Accessed: 8 October 2019).

Beer, D. (2016) *Metric power*. London: Palgrave Macmillan.

Bennett, W. L. and Segerberg, A. (2012) 'The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics', *Information, Communication & Society*, 15(5), pp. 739–768.

Bijker, W. E., Hughes, T. P. and Pinch, T. (eds) (1987) *The social construction of technological systems: New directions in the sociology and history of technology*. Anniversary ed. Cambridge, Mass: MIT Press.

Bijker, W. E. and Law, J. (eds) (1992) *Shaping Technology/Building society: Studies in sociotechnical change*. Cambridge, Mass: MIT Press (Inside technology).

Boden, M. A. (2016) *AI: Its nature and future*. Oxford: Oxford University Press.

Bostrom, N. (2016) *Superintelligence: Paths, dangers, strategies*. Oxford: Oxford University Press.

boyd, danah and Crawford, K. (2012) 'Critical Questions for Big Data', *Information, Communication & Society*, 15(5), pp. 662–679. doi: 10.1080/1369118X.2012.678878.

Callon, M. (1986) 'Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay', in Law, J. (ed.) *Power, Action and Belief: A New Sociology of Knowledge*. London: Routledge & Kegan Paul, pp. 196–233. Available at: http://journals.sagepub.com/doi/10.1111/j.1467-954X.1984.tb00113.x (Accessed: 27 May 2019).

Callon, M. and Law, J. (1997) 'After the Individual in Society: Lessons on Collectivity from Science, Technology and Society', *Canadian Journal of Sociology / Cahiers canadiens de sociologie*, 22(2), pp. 165–182. doi: 10.2307/3341747.

Carey, J. W. (1990) 'Technology as a totem for culture: And a defense of the oral tradition', *American Journalism*, 7(4), pp. 242–251.

Carey, J. W. (2009) *Communication as culture: Essays on media and society*. New York: Routledge.

Castells, M. (1996) *The Rise of the Network Society*. Malden, MA: Blackwell.

Chadwick, A. (2013) *The hybrid media system: Politics and power*. Oxford: Oxford University Press (Oxford studies in digital politics).

Dix, A. J. (2004) *Human-computer interaction*. London: Prentice Hall.

Dung, L. *et al.* (2020) 'Integrating social sciences research with artificial intelligence (AI): A case study from the Great Barrier Reef', *CAUTHE 2020: 20: 20 Vision: New Perspectives on the Diversity of Hospitality, Tourism and Events*, p. 130.

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

Faraj, S., Pachidi, S. and Sayegh, K. (2018) 'Working and organizing in the age of the learning algorithm', *Information and Organization*, 28(1), pp. 62–70. doi: 10.1016/j.infoandorg.2018.02.005.

Frank, M. R. *et al.* (2019) 'The evolution of citation graphs in artificial intelligence research', *Nature Machine Intelligence*, 1(2), pp. 79–85.

Gehl, R. W. and Bakardjieva, M. (eds) (2017) *Socialbots and their friends: Digital media and the automation of sociality*. New York: Routledge.

Gillespie, T. (2014) 'The Relevance of Algorithms', in Gillespie, T., Boczkowski, P. J., and Foot, K. A. (eds) *Media technologies: Essays on communication, materiality, and society*. Cambridge, Mass.: The MIT Press, pp. 167–193.

Govia, L. (2020) 'Coproduction, Ethics and Artificial Intelligence: A Perspective from Cultural Anthropology', *Journal of Digital Social Research*, 2(3), pp. 42–64.

Gupta, S. and Tu, P. H. (2020) *What is artificial intelligence?: a conversation between an ai engineer and a humanities researcher*. Hackensack, NJ: World Scientific.

Guzman, A. L. (2017) 'Making AI safe for humans: A conversation with Siri', in Gehl, R. W. and Bakardjieva, M. (eds) *Socialbots and their friends: Digital media and the automation of sociality*. New York: Routledge, pp. 69–85.

Guzman, A. L. (2018) *Human-Machine Communication: Rethinking communication, technology, and ourselves*. New York: Peter Lang.

Guzman, A. L. and Lewis, S. C. (2019) 'Artificial intelligence and communication: A Human–Machine Communication research agenda', *New Media & Society*. doi: 10.1177/1461444819858691.

Holmström, J. and Robey, D. (2020) 'Materiality and Organizing: Actor-Network Theory Revisited', in Hernes, T. and Czarniawska, B. (eds) *Actor-Network Theory and Organizing*. Lund: Studentlitteratur, pp. 177–201.

Irving, G. and Askell, A. (2019) 'AI Safety Needs Social Scientists', *Distill*, 4(2), p. e14. doi: 10.23915/distill.00014.

Ito, M. (2008) 'Introduction', in Varnelis, K. (ed.) *Networked Publics*. Cambridge, MA: MIT Press, pp. 1–14.

Jouët, J. (2000) 'Retour critique sur la sociologie des usages', *Réseaux. Communication-Technologie-Société*, 18(100), pp. 487–521.

Keller, T. R. and Klinger, U. (2019) 'Social Bots in Election Campaigns: Theoretical, Empirical, and Methodological Implications', *Political Communication*, 36(1), pp. 171–189. doi: 10.1080/10584609.2018.1526238.

Kitchin, R. and Dodge, M. (2011) *Code: Software and everyday life*. Cambridge, Mass.: MIT Press. Available at: http://site.ebrary.com/id/10479192 (Accessed: 31 October 2019).

Lagerkvist, A. (2020) 'Digital Limit Situations: Anticipatory Media Beyond 'The New AI Era'', *Journal of Digital Social Research*, 2(3), pp. 16–41.

Latour, B. (1992) 'Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts', in Bijker, W. E. and Law, J. (eds) *Shaping Technology/Building society: Studies in sociotechnical change*. Cambridge, Mass: MIT Press (Inside technology), pp. 225–258.

Latour, B. (2005) *Reassembling the social: An introduction to actor-network-theory*. Oxford: Oxford University Press.

Latour, B. and Callon, M. (1981) 'Leviathan: How Actors Macro-Structure Reality and How Sociologists Help Them to Do So', in Knorr-Cetina, K. and Cicourel, A. V. (eds) *Advances in Social Theory and Methodology: Toward an Integration of Micro- and Macro-Sociologies*. London: Routledge, pp. 277–303.

Latzko-Toth, G. (2014) 'Users as Co-Designers of Software-Based Media: The Co-Construction of Internet Relay Chat', *Canadian Journal of Communication*, 39(4), pp. 577–595. doi: 10.22230/cjc.2014v39n4a2783.

Lindgren, S. (2014) *Hybrid Media Culture: Sensing place in a world of flows*. London: Routledge.

Lindgren, S. (2020) *Data Theory: Interpretive Sociology and Computational Methods*. Cambridge: Polity.

MacKenzie, D. A. and Wajcman, J. (eds) (1985) *The Social shaping of technology: How the refrigerator got its hum*. Milton Keynes: Open University Press.

Mallein, P. and Toussaint, Y. (1994) 'L'intégration sociale des technologies d'information et de communication: Une sociologie des usages', *Technologies de l'information et société*, 6(4), pp. 315–335.

Manovich, L. (2013) *Software takes command*. New York: Bloomsbury.

Miller, T., Howe, P. and Sonenberg, L. (2017) 'Explainable AI: Beware of Inmates Running the Asylum Or: How I Learnt to Stop Worrying and Love the Social and Behavioural Sciences', *arXiv:1712.00547 [cs]*. Available at: http://arxiv.org/abs/1712.00547 (Accessed: 4 September 2020).

Nardi, B. A. (1995) *Context and consciousness: Activity theory and human-computer interaction*. Cambridge, Mass.: MIT Press.

Neff, G. and Nagy, P. (2016) 'Talking to Bots: Symbiotic Agency and the Case of Tay', *International Journal of Communication*, 10, pp. 4915–4931.

Noble, S. U. (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.

O'Neil, C. (2016) *Weapons of math destruction: How big data increases inequality and threatens democracy*. First edition. New York: Crown.

Orlikowski, W. J. and Gash, D. C. (1994) 'Technological Frames: Making Sense of Information Technology in Organizations', *ACM Trans. Inf. Syst.*, 12(2), pp. 174–207. doi: 10.1145/196734.196745.

Owen, T. (2015) *Disruptive power: The crisis of the state in the digital age*. Oxford: Oxford University Press.

Pinch, T. J. and Bijker, W. E. (1984) 'The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology might Benefit Each Other', *Social Studies of Science*, 14(3), pp. 399–441. doi: 10.1177/030631284014003004.

Pop Stefanija, A. & Pierson, J. (2020) 'Practical AI Transparency: Revealing Datafication and Algorithmic Identities', *Journal of Digital Social Research*, 2(3), pp. 84–125.

Preece, J. (1994) *Human-computer interaction*. Wokingham: Addison-Wesley.

Rahwan, I. *et al.* (2019) 'Machine behaviour', *Nature*, 568(7753), pp. 477–486. doi: 10.1038/s41586-019-1138-y.

Rainie, Harrison. and Wellman, Barry. (2012) *Networked: The New Social Operating System*. Cambridge, Mass.: MIT Press.

Reutter, L. M. (2018) 'Unpacking the Socio-Technological Assemblage of Smart Algorithms - A Case Study on the Production of Machine Learning Algorithms in the Norwegian Labor and Welfare Administration'. Available at: https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2573466 (Accessed: 4 September 2020).

Richardson, Kathleen. (2015) *An anthropology of robots and AI: annihilation anxiety and machines*. New York: Routledge.

Scott, S. V. and Orlikowski, W. J. (2014) 'Entanglements in Practice: Performing Anonymity Through Social Media', *MIS Quarterly*, 38(3), pp. 873–894.

Schatzki, T. R. (2012) 'A Primer on Practices', in Higgs, J. et al. (eds) *Practice-Based Education: Perspectives and strategies*. Rotterdam: SensePublishers, pp. 13–26.

Seidel, S. *et al.* (2020) 'Artificial Intelligence and Video Game Creation: A Framework for the New Logic of Autonomous Design', *Journal of Digital Social Research*, 2(3), pp. 126–157.

Siponen, M. (2004) 'A pragmatic evaluation of the theory of information ethics', *Ethics and Information Technology*, 6(4), pp. 279–290.

Sloane, M. and Moss, E. (2019) 'AI's social sciences deficit', *Nature Machine Intelligence*, 1(8), pp. 330–331. doi: 10.1038/s42256-019-0084-6.

Suchman, L. A. (1987) *Plans and situated actions: The problem of human-machine communication*. Cambridge university press.

Suchman, L. A. (2009) *Human-machine reconfigurations: Plans and situated actions*. Cambridge: Cambridge University Press.

Svensson, J. & Poveda Guillen, O. (2020) 'What is Data and What Can It Be Used For? Key Questions in the Age of Burgeoning Data-Essentialism', *Journal of Digital Social Research*, 2(3), pp. 65–83.

Williams, R. and Edge, D. (1996) 'The social shaping of technology', *Research policy*, 25(6), pp. 865–899.

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

Woolgar, S. and Cooper, G. (1999) 'Do artefacts have ambivalence? Moses' bridges, Winner's bridges and other urban legends', *Social Studies of Science*, 29(3), pp. 433–449.

Woolley, S. (2018) 'The Political Economy of Bots: Theory and Method in the Study of Social Automation', in Kiggins, Ryan. (ed.) *The Political Economy of Robots Prospects for Prosperity and Peace in the Automated 21st Century*. Cham: Springer, pp. 127–155.

JOURNAL OF DIGITAL SOCIAL RESEARCH

WWW.JDSR.IO ISSN: 2003-1998

VOL. 2, NO. 3, 2020, 16–41

DIGITAL LIMIT SITUATIONS: ANTICIPATORY MEDIA BEYOND 'THE NEW AI ERA'

Amanda Lagerkvist*

ABSTRACT

In the present age AI (*artificial intelligence*) emerges as both a medium to and message about (or even *from*) the future, eclipsing all other possible prospects. Discussing how AI succeeds in presenting itself as an arrival on the human horizon at the end times, this theoretical essay scrutinizes the 'inevitability' of AI-driven abstract futures and probes how such imaginaries become living myths, by attending how the technology is embedded in broader appropriations of the future tense. Reclaiming anticipation existentially, by drawing and expanding on the philosophy of Karl Jaspers – and his concept of the *limit situation* – I offer an invitation beyond the prospects and limits of 'the new AI Era' of predictive modelling, exploitation and dataism. I submit that the present moment of technological transformation and of escalating multi-faceted and interrelated global crises, is a *digital limit situation* in which there are entrenched existential and politico-ethical stakes of anticipatory media. Attending to them as a 'future present' (Adam and Groves 2007, 2011), taking responsible action, constitutes our utmost capability and task. The essay concludes that precisely here lies the assignment ahead for pursuing a post-disciplinary, integrative and generative form of Humanities and Social Sciences as a method of hope, that engages AI designers in the pursuit of an inclusive and open future of existential and ecological sustainability.

Keywords: anticipation; existential philosophy; Karl Jaspers; AI; existential media; uncertainty; surveillance capitalism

[*] Uppsala University, Sweden.

The age of great and good actions is past; the present age is the age of anticipation.

Søren Kierkegaard, *Two Ages: The Age of Revolution and the Present Age – A Literary Review*

(30 March 1846, p. 253)

This act of will is my claim to the future tense.

Shoshana Zuboff, *The Age of Surveillance Capitalism* (2019, p. 329)

Give yourself up neither to the past nor to the future.
The important thing is to remain wholly in the present.

Karl Jaspers, *Philosophische Logic* (1958, in *Hannah Arendt/Karl Jaspers Correspondence, 1926-1969*, 1992, p. 153)

1 INTRODUCTION: HORIZONS OF 'THE NEXT CENTURY WITH AI'

AI (*artificial intelligence*) is mounting on the human horizon. Numerous prophesies have in the past few years flooded public discourse, stating that we are inevitably moving into a future driven by autonomous systems with transformative consequences for families and households, the ways we work, produce things, prevent crime, and take care of our vulnerable, sick and elderly (cf. Kelly 2016). For many visionaries, the horizons of AI promise to provide better solutions — increased accuracy, efficiency, cost savings, and speed — to our many problems, and to offer entirely new insights into behavior and cognition. For others, they also usher in new threats and fears about existential risks to our species of an AI superintelligence surpassing that of humanity (Boström 2014). Yet for major agents the main risk seems to be to fall behind in racing toward this new future. Therefore, commercial interests blend with chief geopolitical and military wagers, as exemplified by North American stakeholders who aim to ensure that "the coming AI century is an American one".[1]

Boosted by corporate concerns about avoiding another 'AI Winter' – when confidence in the promises and potentials of these technologies may languish and investors may withdraw – the ethical imperatives raised by these technologies have also spurred an entire 'industry' which mobilizes for example investors, academia, governments, engineers and think tanks seeking to promote and secure sustainable, benevolent, responsible and ethical AI. Yet, positing AI as both a medium to and message about (or even *from*) the future, measured as well as unbridled responses, utopian as well as dystopian scenarios, in fact allow this technology to eclipse all other possible prospects (cf. Dencik 2018, 2020, McQuillan 2019). The expectations for 'the next AI century' are here saturated with what Donald MacKenzie and Judy Wajcman (1999) call a "technological trajectory," which is an

[1] The Center for New American Security promises to ensure "…a new technological era where America's national security—and that of U.S. allies and partners—is more secure, its economy is poised to flourish, and its norms and values underpin AI technologies worldwide" https://www.cnas.org/publications/reports/the-american-ai-century-a-blueprint-for-action. The Chinese government and weapons industry, on their part, foresee that lethal autonomous weapons will be commonplace by 2025, and claim that ever-increasing military use of AI is "inevitable […] We are sure about the direction and that this is the future." Gregory C. Allen reports for The Center for New American Security about the Chinese AI policy, here citing Zeng Yi, a senior executive at China's third largest defense company, Norinco, at the Xiangshan Forum. See https://www.cnas.org/publications/reports/understanding-chinas-ai-strategy. See also China State Council, "Made in China 2025," July 7, 2015; English translation available at http://www.cittadellascienza.it/cina/wp-content/uploads/2017/02/IoT-ONE-Made-in-China-2025.pdf

institutional form of technological change that entails a "course of development that seems natural and autonomous" (Gates 2011, p. 24). The massive mobilization of this future across the board is thus awash with "illusions of inevitability" (ibid), that is what Shoshana Zuboff in *The Age of Surveillance Capitalism*, recently calls 'inevitabilism' (2019, p. 194, pp. 222-224). This future is now, as Zuboff alerts us to, part of a larger project of instrumentarian and rogue surveillance capitalism, which has powerfully lured us all into an iron cage of datafication where human experience is rendered as behavioral data. This implies a massive mining of our bodies and inmost lives, excavating the depths of human existential needs, without consent.

Zuboff argues that this is a new frontier of power, a new form of capitalism which operates through a 'ubiquitous apparatus' (that is Google, Facebook, Apple etc.) that declares the right to harvest our behavioral data and to shape behavior in the real world. This apparatus has hijacked the promises held by new media technologies and digitalization. It thus succeeds primarily by exploiting what second modernity humans caught up on the grids of callous bureaucracies, actually crave and expect of life: their inner sense of worth and dignity, their search for value, meaning and self-expression. In the process of filling those vast voids with effective, accessible technologies that promise to make life worth living, absolute *certainty* has replaced trust for the purpose of control. This for the ultimate benefit of the few and with nothing less than the human future in the balance. Beyond what she calls 'the prediction imperative' (ibid, pp. 197-200) the tech agents are within the 'economies of action' involved in molding our future behavior, and thereby rob us of a future tense. Hence, AI – one key technology in this drama – not only sits on but seemingly also closes the horizons of futurity.

This enclosing scenario might make Danish philosopher Søren Kierkegaard (1813-1855) roll over in his grave. In his fervent critique of 'the present age,' (1846) he painted it as devoid of passion; serious, abstract and calculating while indulging itself in endless publicity and public relations activities, only offering 'reflection' in the shape of sober thought or bland imagery. He argued that in the present age of modernity, we are reduced to quantifiable common denominators – to a 'public' – and in fact disabled from *real action.* Nothing is unforeseen: "The age of great and good actions is past; the present age is the age of anticipation" (ibid, p. 253). 'Anticipation' for Kierkegaard thus refers to the urge of exacting everything in advance, which also feeds into the leveling of the value of the unique singular human being, and in turn disables and nullifies human choice, action, and ethical responsibility.

The horizons of AI are one evident outcome of the statistical attitude that Kierkegaard deeply lamented in his time. For contemporary technoprogressivists who promise to leverage AI to solve humanity's many problems, 'anticipation' is understood in ways that reflect how modernity at large executes "an *'abstract future'* subject to deterministic or probabilistic laws for science, economics, and public administration" which in turn leads to "the pursuit of *empty futures*" (Adam and Groves 2011, p. 17, italics added). The hype around predictive AI is thus forging such a rampant form of modernity which entails a "de-contextualized future emptied of content" […] "open to exploration and exploitation, calculation and control" (Adam and Groves 2007, p. 2).

The purpose of this essay is to scrutinize the 'inevitability' of AI-driven abstract futures, and probe how such imaginaries become living myths, by attending how the technology is embedded in broader appropriations of the future tense. In addition, I suggest that we in a creative and unorthodox manner turn to the philosophy of German existentialist Karl Jaspers in order to provide an existentialist understanding of (media) futures and of anticipatory media. I see anticipation as a centrally important concept to reclaim and safeguard from those less good forces who own it now (the robber barons of the platform society, the high tech monopolizers) but also ultimately for scholars of digital society and of existential media studies to set out to collaboratively theorize. This is because imagining the future is an *existential practice* (Josephides 2014), an irreducible aspect of being human that belongs to us and to our faculty of anticipation.[2] I take my cue from Barbara Adam and Chris Groves who "imagine different ways of acting responsibly in creating futures." Through a Heideggerian framework of *care* they offer "some new conceptual coordinates for thinking about the ethical underpinnings for our relationship with the future and for reshaping the legal and thereby the political expressions of our responsibilities to it. They might help restore a sense that the future matters" (2011, p. 17-18).

I suggest that Jaspers' thinking will offer precisely such "new conceptual coordinates", that can be helpful in this project of conceptualizing anticipation and a lived future in and of the present, since it will forefront the inherent uncertainties of being and what Jaspers calls the *limit situations* of life (1932/1970). The present age of technological transformation and of escalating multi-faceted and interrelated global crises (Gasper 2018) – I argue, is a *digital limit situation* in which there are entrenched existential and politico-ethical stakes of anticipatory media. Attending to them, taking responsible action, constitutes our utmost capability and task. In fact, responsibility is the cornerstone of Jaspers'

[2] Scholars in the field of anticipation studies see anticipation as a faculty fundamental for both human flourishing, creativity, ethics, politics and for society as a whole, and for the technologies we build and embrace to ultimately enable (cf. de Miranda et al 2016).

political philosophy and “ethical theory, which sees human life as self-creating, autonomous and plural, but also supremely, if not universally, accountable” (Thornhill 2002, p. 6). Taking responsibility for AI also means, importantly, pausing in the present in order to collaboratively shape the future. Yet, this feeling that we are *on the brink of something*, the sense of both occasion and urgency, is in fact also a ‘zeitgeist’ of sorts, a major current of our time resonating in popular discourse as well as formal, academic and economic thinking (Guyer 2016). These gravitations to the present moment now enhanced by the pandemic, compel a reorientation: a slowing down to think about core values and chief priorities – both in life, scholarship and society (cf. for example Corpus Ong and Negra 2020, Henderson 2020, Dencik 2020). The present moment is in fact a time when human beings – in all our diversity – could potentially begin to realize what Barbara Adam and Chris Groves call a *concrete practical future* with technology (2007). AI on the human horizon thereby presents us with a momentous assignment. How we respond depends on how we conceive of media.

2 FROM LIFE-APPARATUS TO EXISTENTIAL MEDIA

In a classic move, Zuboff opposes the “tyranny of prediction” to a “human future”. This reflects a distinction in Jaspers’ work (as well as among many critics of modernity and mass culture of his generation), between technological deprivation and human value (1931, 1951). Echoing Kierkegaard, Jaspers sees a problematic hollowing out of meaning and value — the result of modern technological culture in which “[e]ssential humanity is reduced to the general” (Jaspers 1931, p. 49). For Jaspers this has wide consequences for limiting humanity: “Limits are imposed upon the life-order by a specifically modern conflict. The mass-order brings into being a universal *life-apparatus*, which proves destructive to the world of a truly human life” (ibid, p. 44, italics added). He further states that “[t] he universalization of the life-order threatens to reduce the life of the real man (sic!) in a real world to mere functioning” (ibid, p. 45).

To renew its relevance, Jaspers’ systemic critique will obviously need an upgrading.[3] For example, important debates in critical data studies have problematized not only how technological systems are exploiting our datafied lives, but also how they are rehearsing and amplifying, instead of checking, human prejudice, bias and stereotyping (see for example Noble 2018, Bucher 2018, Eubanks 2018). It is also necessary to incorporate contemporary empirical insights from media sociology and anthropology,

[3] This is a key methodological approach in existential media studies, which in interesting ways overlaps with calls for upgrading theoretical paradigms to the actualities of a data-driven social world (Lindgren 2020).

inspired by (post-)phenomenological and new materialist understandings of the onto-epistemological dimensions of human-data assemblages. Big Data and biometric technologies for example are both part of the body politic and habitual, meaningful, entangled and mundane data with varied and contextually bound uses and meanings. Even if they exploit, surveil and reduce humans as Jaspers would say, they are also productive as they bring into being new forms of knowledge and social relations, new assemblages and webs of everyday ordinary life-flow, new data subjectivities and forms of embodiment (see for example Lupton 2016, Pink and Fors 2017, Pink et al 2017, Kennedy and Hill 2018, Guzman 2019).

In keeping with such acumens, I however take my main lead from Jaspers in placing particular emphasis on limits – as well as how they relate to radical uncertainty, openness and fecundity in the present – to offer nothing more and nothing less than what I believe to be central prompts for thinking about an existentially sustainable future in which we become human with machines (cf. Kember and Zylinska 2011). This themed issue seeks to shed light on the fact that AI is always socially embedded. I argue in addition that precisely because humans and machines are co-implied and co-constituted recursively and because data are mundane and deeply enmeshed in our lives – and in light of critical insights about surveillance capitalism – the question of how to realize *existentially sustainable anticipatory media* is even more pertinent to raise. I will offer a twin reconceptualization of anticipatory AI as existential media, and of existential media as in fact anticipatory by nature. Hence, as an exercise in existential media studies, which combines a materialist understanding of media with Kierkegaardian and Jaspersian wisdom, I submit that existential media (Lagerkvist 2016) – that both condition and are conditioned by the digital limit situation – have four interrelated properties that I hope to substantiate throughout. They are, first as John D. Peters would say "our infrastructures of being" (2015, p. 15) which means that they ground us materially in existence. Yet, they also, second, throw us up into the air, and in their contingency they in fact ambivalently limit us and offer radical openness at the same time. Third, they furthermore speak to and about originary human (yet unevenly distributed) vulnerability and deep relationality. Finally, they demand responsive action. The latter property is heavily influenced by the existential stakes of the present age of anticipation as prediction.

3 EXISTENTIAL STAKES AND SITUATED BEING(S) OF AND BEYOND DATA

Indeed, the human capacity to anticipate, aspire, and look forward – what Edmund Husserl calls 'protention' – seems kidnapped by machines and screens (Lagerkvist 2018). For Zuboff, the entire future for humanity is therefore now at risk. When the future tense itself seems lost, there are deep existential stakes. Zuboff is passionately searching for an existential language to describe this sense of demise and loss of possibilities for willing the future itself, in a world of all-pervasive datafication and automation. Echoing one influential strand of the existentialist tradition which submits that the very possibility of projecting ourselves into a future (Heidegger 1927, Sartre 1943, de Beauvoir, 1947, Schutz 1972, Arendt 1978) is key for what makes us human, she holds that "the freedom of will is the existential bone structure that carries the moral flesh of every promise, and my insistence on its integrity is not an indulgence in nostalgia or a random privileging of the pre-digital human story as somehow more truly human" (2019, pp. 330-331). She further contemplates:

> No matter how much is taken from me, this inward freedom to create meaning remains my ultimate sanctuary. Jean-Paul Sartre writes that 'freedom is nothing but the *existence* of our will,' and he elaborates: 'Actually it is not enough to will: it is necessary to will to will.' The rising up of *the will to will*, is the inner act that secures us as autonomous beings who project choice into the world and who exercise the qualities of self-determining moral judgment that are civilization's necessary and final bulwark. (Zuboff, 2019, p. 290, italics in original)

Hence, we may ask in similar vein whether Big Data, AI and machine learning of the present age, with their technocratic, entrepreneurial and capitalistic ethos, will further hamper (as Zuboff details) the prospects for realizing ourselves through projects of our will. Or will they even relieve humans of the responsibility they have for their lives, for each other, and for the planet? Do they in fact offer an escape from that responsibility for those Kierkegaardian choices and actions that shape the future? Yet, while Zuboff's freedom of will is important, I hold that we actually need an even broader existentialist purview to address the existential stakes of AI futures and their imaginaries. We thus need to ask in addition whether these technologies could in fact become part of what Arjun Appadurai describes as an *ethics of possibility* based on "those ways of thinking, feeling and acting that increase the horizon of hope, that expand the field of the imagination, that produce greater equity" within our aspirational capacities so as to "widen the field of informed, creative and critical citizenship"? (2013, p. 295).

In their seminal work in anticipation studies, Barbara Adam and Chris Groves have identified a weakness within the abstract futures model: "the key problem for an empty futures perspective remains that the future is not simply beyond the present but is a latent and 'living future' *within* it" (2011, p. 17, italics in original). They argue for turning to the existentialist tradition to reconceive of the living future, which we have to tend to and care for, by caring for each other, as well as for the objects, phenomena and progressive ideas, and other beings that we share our existence with (ibid, p. 24). They conclude that different forms of social action "facilitated by advanced technologies and complex social structures need to be based around a different image of the future" (ibid, p. 17). One possibility, they hold, is the kind of "'lived future' that is articulated in Heidegger's (1998) account of Dasein's characteristic temporality" in combination with perspectives from Hans Jonas' biology. They hold that "[t]he perspective of a lived future, dependent on a situated subject whose being is an issue for it, relates itself very differently to the living, latent futures of action that surround it and in which it itself is embedded" (ibid, p. 18).

As discussed above, consulting Jaspers enriches and complements the temporal subjectivities of for example Heidegger's sense-making and resolute, yet anti-subjectivist, *Dasein* and Sartre's subject that wills to will. For Jaspers there are three modes of being human. The first is *empirical existence* existing in a material world of basic desires. Second, we are *consciousness in general* which pertains to the faculty of abstract thinking, logos and mathematics. Third, human beings are *spirit* which encapsulates our attempts to create a whole, a world view, out of fragments in for example ideologies and religions. But there is yet one form of potential being: as realized *Existenz*. This form defies objectivity: it defines human beings in authenticity, singularity, inwardness and transcendence – and in truth in/as communication. Realized *Existenz* is a potential for each of us, but also something we may fail to be.

In Jaspers' philosophy human beings furthermore always and inevitably find themselves in situations: "existence means to be in a situation" (1932/1970, p. 178). There are two types of situations. The first is the immanent type of *situations in existence*. In general, we are born into a particular time and space, in which we face and share certain historical circumstances and conditions. Our being in situations in existence is also concrete, every day, material. This applies to us all, yet situations in existence are socially diversified. This type of situatedness is "*a reality for an existing subject who has a stake in it*, a subject either confined or given leeway by the situation in which other subjects, their interests, their sociological power relations, and their combinations or chances of the moment all play their parts" (1932/1970, p. 177, italics in original). This empirical existence can be captured by *data*:

> At each moment I exist by given data, and I face given data to which my will and my actions refer. This is how I am for myself as empirical existence, and how the definite world to which I have access exists for me as a datum I can mold. The real situation confines me, by its resistance, limits my freedom and ties me to restricted possibilities. (ibid, p. 185)

But there is infinitely more to being human in our situation than our data – or perhaps as we would today put it, our 'metadata'. There are also the *transcendent limit situations of life:* "Situations like the following: that I am always in situations; that I cannot live without struggling and suffering; that I cannot avoid guilt; that I must die – these are what I call boundary situations" (ibid, p. 178). Limit situations of for example crisis, conflict and death, underscore the singularity of our human lives. We have to enter into them with open eyes; they require something of us, and offer a possibility of realizing our *Existenz* together (1932/1970, p. 64). While the limit situation affords an important role to inwardness it is both a shared affair, and tied to political and social responsibility. As Chris Thornhill has pointed out, "Jaspers' theory of existential interiority is in fact at all times correlated with a strong Kantian and Weberian dimension, which views existential authenticity as the foundation for an ethic of social and political responsibility, not as the static celebration of isolated subjectivity" (2002, p. 3). This is why the concept of the digital limit situation is apposite for describing this uncertain moment which simultaneously entails a future seemingly destined to be forged by AI; a present before which we are called to awaken ourselves collectively. The concept grasps the urgency and severity of those cataclysmic transformational forces of the present moment; it allows for thinking about the gravity of the situation and the responsibility we have for it.

Drawing inspiration from, yet expanding on Jaspers' thinking I have reconceived of humans (and of 'media users') as singular-plural, deeply relational, technological, situated, embodied and responsible beings – as *coexisters* (Lagerkvist 2016, 2019). Contingent upon limits of both knowledge and self-awareness, they exist within the biosphere together with other humans, machines and more-than-humans. The coexister is not the discrete rational and moral subject of old-school humanism who is certain, independent and disembodied. Instead the coexister is that being that strives, hurts and hopes and is often clueless; that realizable *Existenz,* who possesses the human potential for flourishing which we always do in deep relationality with both fellow humans, as well as with animals, tools, machines and networks. Coexisters are *thrown* into the contemporary digital limit situation; deeply entangled they still possess the capacity to act and chose and respond – and anticipate – yet within limits and never in isolation. In that way coexisters are in fact proficient to collaboratively chart a (media) future in carefully attending to the present.

Here AI technologies and imaginaries play major roles, bearing on how we may or may not anticipate the future. In order to further open these vistas, I will offer a minor mapping of key concepts, definitions and insights within *anticipation studies.* How do contemporary media futures map onto the concept of anticipation itself? And what are the alternatives – how can we conceive of anticipation existentially?

4 ANTICIPATORY MEDIA: FROM ABSTRACT MEDIA FUTURES TO ANTICIPATION PROPER

Media studies as a field has a peculiar and complicit relationship to media futures and their imaginaries. Media of the bleeding edge figure more or less unconsciously, as both pointers to and foretellings about 'the Future'. Due to the anticipatory features of data and predictive modelling, however, the relationship between media and the future is changing. This has in turn prompted a tide of explorations of the future tense in media studies (see for example Andrejevic 2019, Hong and Szpunar 2019, Zylinska 2020, Pentzold, et al 2020), to which I also hope to contribute.

AI is *anticipatory media* in several senses. The phenomena we call AI seem to be, both as a set of media technologies and an analytic phenomenon, essentially *about anticipation.* They materially and symbolically foresee and thereby bring a world into being. AI forecasting, modelling, prediction, and prognosis advises, predicts, if not always outright decides, "about how data should be interpreted and what actions should be taken as a result" (Mittelstadt et al 2016, n.p.). As Christian Pentzold (et al) recently put it: "Digital media, networked services and aggregated data are beacons of the future" (2020, p. 2). Hence, they "do not only forecast uncharted times or predict what comes next," they are, it seems, "both prognostic and progressive media: they don't await the times to come but realize the utopian as well as dystopian visions which they have always already foreseen" (ibid, p. 7). AI thus co-creates the future in predicting it.

Coupled with the ideology of dataism, such aptitudes of AI thereby seemingly assume metaphysical, magical or even divinatory capacities to foresee the future (van Dijck 2014, Chun 2016, Esposito 2018). As Joanna Zylinska maintains, these technological imaginaries also belong to a narrative with a gendered tenor of "messianic-apocalyptic undertones" and "masculinist-solutionist ambitions" (2018, p. 15). Hence, the advent of this technology is in the guise of anticipatory media that may *salvage* us. This furthermore feeds into Jane Guyer's analysis of contemporary temporalities (2007, 2019) in which the *near future* – a social and material world that we could previously imagine, plan, hope for and intelligibly try to shape and realize – has disappeared. This has been replaced by the combination of an absolute sense of *the next moment* – a punctuated time of rigid calendrics and dates modelled upon the finance sector – with *the long-term,* widely touted both in the myths of macroeconomics of eternal progress, and in evangelical ideas of prophetic time. AI thus arrives on the empty horizon of the future, and both fills up that next moment with datafied answers, and fulfills the expectation of an arrival; a salvation at the end times. In fact, the notion of *the next century with AI* is itself downright illustrative of this hybrid temporal modality of the 'next' and the 'infinite'.

This form of future orientation goes in the field of anticipation studies under the name of *forecasting* (Poli 2017, p. 67). Forecasting focuses on capturing continuity through quantitative models and "is the properly predictive component of futures study. Its models tend to adopt either a very short – as with econometric models – or a very long – as with climate change models – temporal window" (ibid); hence a combination of the next and the infinite. As already noted, Barbara Adam and Chris Groves distinguish between two types of futures: *abstract* and *concrete futures*. "Abstract futures […] correspond to forecasting extrapolations, or more generally to system dynamics modelling in which the future is seen as a projection and a product of the past" (Poli 2017, p. 34). Such "*present futures*" are "imagined, planned, projected, and produced *in* and *for* the present" (Adam and Groves 2007, p. 28, italics in original). These are for example economic and scientific forecasts that colonize the future from the present through derivatory models of exploiting the future for gain (Miller 2007, Halpern 2018). As discussed above, Zuboff has pinpointed the latest and most pervasive of all such exploits of the future though forecasting. In this diagnosis, the future has thus returned, via anticipatory media, which seem to have kidnapped it at once.

To theoretically and imaginatively propose existentialist openings, one must first possess a more fine-grained concept of anticipation. The field of anticipation studies further distinguishes between *forecast*, *foresight* and *anticipation* (Poli 2017, p. 67). While forecasting implies prediction and calculus, *foresighting,* by contrast, is not predictive. It produces a variety of possible futures to challenge the mindset of decision makers. It is qualitative and focuses instead on discontinuities. *Anticipation,* in turn, involves both a future oriented attitude and using the knowledge one has gained from that attitude to plan and act accordingly (ibid, p. 35). Hence, a system behaving in an anticipatory manner takes decisions in the present according to anticipations about something that may occur in the future.[4] Using the future is in fact the very meaning of 'anticipatory behavior'. It seems then that AI is anticipatory if this is the main qualifying characteristic.[5]

[4] The field of anticipation studies thus furthermore differentiates between *anticipation* and *anticipatory system*. An anticipatory system is defined as a system "containing a predictive model of itself and/or its environment which allows the system to change state at one instant in accord with the model's predictions pertaining to a later instant" (Rosen 1985/2012 in Poli 2017, p. 2).

[5] As argued by Rovatsos, AI displays in line with Poli's analysis "elements of an anticipatory process: A model of the system is used to consider different alternatives about

Yet, anticipation also shares some features with foresight: it is non-predictive, qualitative, complex and focused on discontinuity and uncertainty. Hence *anticipation proper* also has an impredicative nature. Roberto Poli traces this to for example aspects of biology and society that fail, or refuse, to be reduced to quantification. For example, within the study of autopoietic systems and within relational biology, there is an acknowledgement that all the dynamic processes within an organism are self-referential and mutually linked. Poli explains: "The thesis of impredicativity has wide consequences, one of the most important being that all the information describing an organism will never be completely captured by any algorithmic (i.e. mechanistic) model" (ibid, p. 19). In discussing anthropological perspectives on anticipation, he concludes that theological reflections on the future are, perhaps surprisingly, "in perfect accord with the theory of complex and impredicative systems" (ibid, p. 28). The exegetic tradition thus similarly concludes that:

> The real future is 'uncertain' and is not just the unfolding of our present ideas or strategies. It is not simply a calculated human creation involving 'plans plus time.' Rather the open future that comes to meet us brings surprises. That unforeseen future requires provisionality, since it cannot be calculated or controlled. (Prusak cited in Poli 2017, ibid)

Hence, by these criteria 'anticipatory AI' would in fact flunk to be an example of anticipation proper which shares qualities with the limit situation – in particular that of *uncertainty*.

5 UNCERTAINTY: THE NECESSARY HABITAT OF THE LIVED FUTURE PRESENT

With support from anticipation studies, we can actually establish that the real future is uncertain thus containing uncontrollable and incalculable openness. There is something liberatory about straightforwardly proclaiming that the future is existential in this way.[6] The anticipatory dynamic itself – understood in terms of the above-discussed features of anticipation proper, which includes the capacity to keep futures radically open – is thus integral to the limit situation. And concomitantly, as

what might occur in the future and makes decisions about what action to take in the present. And, the future is seen as a projection of the past through the present" (2019, p. 1508).

[6] As we have seen, the future itself has an open-ended, ambivalent and deeply existential quality. Indeed, the ambivalence of the future is profoundly true both when it is sought through a forward-looking attitude (in a practical lived sense, in a latent future in the making), and when it is pursued as project and projection (as a plannable, pre-given and 'abstract future').

coexisters we are in fact beings of deep uncertainty assigned to navigate, anticipate and thereby pursue a lived future in attending to what is called upon us within the limits of the present: within the digital limit situation. We are thus inevitably involved in what Adam and Groves call the latent future: our dealings and doings, our media practices and projects, our designing and deliberating – including careful academic and philosophical thinking in and about the present age – all in fact constitute *futures present* why they are of import and of consequence.

Jaspers' philosophy delves into both presentness and uncertainty in creative ways, since it sits on the limits of the known and the controllable. His approach allows us to recognize that carefully attending to the present situation constitutes the core of what makes us human. And thus, possibly the core of realizing a sustainable, concrete future with media. In the concluding chapter of *The Perennial Scope of Philosophy* entitled "The Philosophy of the Future" Jaspers offers an understanding of truth in time, as belonging ultimately to the present:

> But is life for the future the essential import of our work? I do not believe so. For we serve the future only in so far as we realize the present. We must not expect the authentic only from the future. Even though this presentness cannot in fact attain to durable consummation, in which I can rest and endure in time, it is nevertheless possible in penetrating this actuality to penetrate in a sense the eternal present in its temporal manifestation. The actuality of the truth in time is, to be sure, as impossible to capture as an optical image, – but it is always with us. (1949, p. 157)

He argues that a philosophy of the future must be able to take hold of the riches and possibilities of the present, in which we can realize ourselves as what he calls living *Existenz* with other *Existenz*. In asserting similarly a future present, Chris Groves echoes Jaspers in insisting on a concrete, embedded, relational and existential future: "What presence does the future have, here and now, and in what way does our relation to it affect our wellbeing and capacity for flourishing? Not any specific future, but the future as an existential dimension of our relationship to others, to ourselves and to the world" (Groves forthcoming, n.p.).

This emphasis on the future present thus resonates with the limit situation, which if seized authentically and sincerely, can be a site for opening new futures. Importantly for my argument in the following, the human limit situation is indeterminate and never fully surveyable. Uncertainty is thus key. Shoshana Zuboff relies on Hanna Arendt's concept of will as "the organ for the future". "The power of will", Zuboff argues following Arendt, lies in "its unique ability to deal with things",

> 'visibles and invisibles' that have never existed at all. Just as the past always presents itself to the mind in the guise of certainty, *the future's main*

> *characteristic is its basic uncertainty,* no matter how high degree of probability a prediction may attain. (Arendt 1978, cited in Zuboff ibid, p. 329-330, italics added)

As Zuboff maintains, the most foundational aspects of human existence are today embezzled by surveillance capitalism, with the ultimate goal to combat 'chaos'. But, as she acknowledges, "uncertainty is not chaos but rather *the necessary habitat of the present tense….*" (2019, p. 336, italics added).

In the existentialist tradition freedom and necessity/finitude – corresponding to uncertainty and situatedness, openness and limits (see de Beauvoir 1946, cf. Withy 2011) – are fundamental and irreducibly interdependent dimensions of human existence. Uncertainty and unhomeliness (as much as freedom) thus belong to the human condition itself. They can also, by contrast, be seen as a dimension of contemporary and historically specific times of political, ecological, epidemiological and technological crises with asymmetrical consequences for those affected (cf. Akama et al, 2018, p. 19). Guyer ponders similarly:

> One could perhaps reduce all this to an ahistorical 'life in uncertain times' or an ancient philosophy of risk 'taken on the flood' (to quote Cassius in Shakespeare's Julius Caesar). There is, however, a *historical specificity to uncertainty now*. It is an emerging chronotope … honed into technologies that can deliberately unsettle and create arbitrage opportunities and gridlocks as well as logistical feats of extraordinary precision and power. (Guyer 2007, p. 418, italics added)

The latter reflects Zuboff's prediction imperative, and it describes the quest for complete certainty within surveillance capitalism. In Zuboff's own words, which again brings what I call the digital limit situation to mind:

> I suggest that we now face a moment in history when the elemental right to the future tense is endangered by a panvasive digital architecture of behavior modification owned and operated by surveillance capital, necessitated by its economic imperatives, and driven by its laws of motion, all for the sake of its guaranteed outcomes. (2019, p. 331)

AI as anticipatory media in this reading, will offer nothing but guaranteed prediction, and in blackboxing its own workings, surveillance capitalism may further increase uncertainty (ibid, pp. 342-343).

I see uncertainty as a perennial dimension, belonging to the human condition – to being itself – even as we are simultaneously situated differently in political and social terms, which deeply affect our lives. The technologically enforced lifeworld may however usher in heightened uncertainties, vulnerabilities and existential anxieties (Lagerkvist 2016, 2019, see also Adam and Groves 2007, p. 55). I thus combine conceiving of uncertainty as a given and as contextually dependent, and of vulnerability as ontological and social (MacKenzie et al 2014) and in effect as socio-technological at the same time. In line with how a number of scholars are today arguing for embracing uncertainty, I hold that it should be subjected to new forms of post-disciplinary scrutiny (Akama et al 2018, Halpern 2018, Guyer 2019). This move is necessary to take on, both conceptually and practically, if we aim to contribute to not only how we understand the future with media, but to how we actually intervene imaginatively in its making.

6 COMPLICATING MATTERS AND METHODS OF HOPE

How do we dissolve the spell of the horizons of the 'new AI era' and bring about alternatives? How do we act and "think what we are doing" (Arendt 1958, p. 5) in the present moment? By pausing (which is in the very nature of the limit situation!) we will note a cluster of complicating matters. First, in a disturbing manner the aforementioned colonization of anticipation for profit, also applies to the 'uncertainties' of being. Jane Guyer illustrates how the language of 'brinks' and 'adventures', 'emergencies' and 'indeterminacies', have filled the evacuated near future, both in popular and formal discourse as well as in economic thinking and academic debate (Guyer 2016). And in 'the present moment' the limit situation seems apprehended in AI projects such as "AI for Earth" or "AI for Good" at Microsoft,[7] or in the technologies launched for tracking contagion during the current Covid-19 crisis (Klein 2020). It is not far-fetched to suggest that the tech agents are seizing their opportunity. Boosted by a righteous project framed within well-meaning goals and benign intentions of salvaging the planet and the species, they are operating through the logic of surveillance capitalism at the same time and take their imperatives of mining the depths of our lives even farther. The digital-human limit situation is ultimately in the hands of very powerful agents, with a gargantuan apparatus of rhetorical and infrastructural means at their disposal.

Hence it seems that it is not enough to reclaim the future tense; it is also urgent to lay claims anew to the very limit situation itself and meticulously ruminate on its meanings and stakes. This implies an awakening. As Jaspers puts it: "Awaking to myself, in my situation, I raised the question of being" (1932/1969, p. 45). In fact for Jaspers: "[p]hilosophizing starts with *our situation*" (ibid, p 43, italics added). This means to raise the most profound philosophical questions – together – in search for new light ahead: What is the meaning of our technologized existence? How do we wish to live our lives together on the planet with machines? How can we diversify AI-driven lifeworlds? Can 'autonomous

[7] https://www.microsoft.com/en-us/ai/ai-for-good

systems' be subject to a democratic screening, a vetting, so as to guarantee a bedrock of non-negotiable goals – perhaps justice, equity, sustainability, non-violence[8]. And how does automation entangled with human needs and necessities change our 'situation'? How can these technologies be harnessed for realizing an existentially and environmentally sustainable and *concrete* future which is "embedded, embodied and contextual" (Adam and Groves 2007, p. 11)? Could they in fact be "technologies of the imagination" (Sneath et al 2009) that generate something beyond the ethos of surveillance capitalists?

Time has come, as many seem to agree 'in this moment' to re-center concerns and agendas and to in fact reclaim a more utopian future. In this spirit, Joanna Zyliska follows Franco 'Bifo' Berardi in raising the questions about whether our future has already been expended or whether it can still be redeemed. Drawing on his idea there is a multiplicity of immanent possible futures (Berardi 2017) and invoking something close to what I call the digital limit situation, Zylinska argues:

> The present moment, with its ecological and economic destructions, and the material and discursive havoc wreaked upon our planet, seems to suggest humanity is on a downward trajectory, that it has already ordered in its own expiration. Yet, contrary to the predictions of the various fetishists of the apocalypse, I want to follow Bifo in arguing that our shared future has not yet been totally spent, irrevocably conquered or deterministically designed. And so, amidst the ruin of our current political thought, a possibility of another, more utopian, future can perhaps be sought and fought for. (2020, p. 148)

Enter *hope,* which is importantly not a thing, a possession: it is a "method for self-knowledge" (Miyazaki 2004, p. 139), allowing for a re-orientation of oneself and of knowledge toward the future (cf. Kavedzija 2016, p. 4). The method, used by the disenfranchised Suvavou people, resonates with the

[8] Indeed, this is already an ongoing endeavor as for example when "The New AI Alliance" is inviting the citizens of Europe into a dialogue on AI applications and ethics. As they put it in their mission statement: "To lay the foundations of responsible development, this platform will host a dialogue on the principles that should govern our technological future and on their practical implementation. A High-level Expert group nominated by the European Commission will engage the members of the Alliance in the discussion. […] I would like to invite you to reflect on what the future holds for all of us and how we can best prepare for it. Let us use the European AI Alliance to shape our digital future together. I hope you will take this opportunity to actively participate in the debate!" (Lucilla Scolli, The New AI Alliance, EU, June 13, 2018). The intention in this essay is to argue for the need to begin this discussion in an existentialist manner, beyond instrumentarian deadlocks and technocratic assumptions, and in deep acknowledgement of the fact that how we define human existence affects how we may take on our task to care for the future present.

limit situation: "the moment of hope that emerged at the moment of abeyance of agency was, then, simultaneously open and closed" (Miyazaki 2004, p. 106). Similarly, seizing the limit situation "allows for the possibility of an uncertain future" (Jaspers 1932/1970, pp. 183-184). Uncertainty is, as already discussed, the flip side of existential freedom: "The unrest in this boundary situation is that what is up to me lies still ahead" (ibid). In full recognition of both limits, suffering and exposure in the limit situation, Jaspers still argues that "[it] is possible for a more profound serenity to rest on grounds of inextinguishable pain" (ibid, p. 195). Uncertainty may thus be generative (cf. Akama et al 2018, p. 45).

In ways that echo these insights, and much in line with how I read Jaspers, Marianne Hirsch launches the notion of *vulnerable time*, to ultimately argue that "unlike trauma, vulnerability shapes an open-ended temporality – that of the threshold of an alternate, reimagined reality" (Hirsch 2016, p. 80). I hold that digital-human vulnerability is situated on this very threshold – and that it can produce self-knowledge for networked humanity. As a method of hope, Zuboff is in favor of replacing the abstract future of the surveillance capitalists, with a plan of her own for third modernity humans. She suggests that instead of an individualistic framework of counter-declarations of hiding from the networks, we need synthetic declarations involving civil society, collective action and legislation (2019, p. 344). We must will to will together! Zuboff is terrified of the companies taking their 'responsibility,' as this consequentially becomes part of their logic for extraction and prediction. Zylinska also argues against the CSR of 'ethical AI' which she sees as a way for companies to try and suspend, and ward off, policy intervention (2020, pp. 34-35). Mark Andrejevic sees risks in offloading human agency and judgement to machines and wants us to move beyond the "ethical turn" and replace it with "data civics" (2020). The emphasis should thus be placed on a veteran method of hope: the modernist form of near-future planning which should imply regulating politically and legally the leeway, scope and scale of the current tech giants, and thereby controlling their development of AI in the service of humanity. We may note that by similar token for Jaspers, the general situatedness of human life encompasses change within, it is in essence transformable (Jaspers 1932/1970, p. 178). From this perspective even a future seemingly encapsulated by prediction technologies belongs to this quality of the situation:

> I have to put up with them as given, but not as definitely given: there remains a chance of transforming them, even in the sense that I can calculate and bring about situations, in which I am going to act as given henceforth. This is the character of purposive arrangements. In technological, legal, political action *we create situations*: We do not proceed directly toward a goal, we bring about the situation from which it will arise. (ibid, italics in original)

For Jaspers, in his relentlessly hopeful manner, these modernist plans thus also contain openings. An alternative to 'the New AI Era' would be to envision regulated and controllable AI in the hands of human collectives as aids in the mundane and deeply existential projects of sustaining relationships to each other and to our planet. In order to bring about a century of care and attendance, as Jaspers would probably suggest in his insistence upon limits, the wise thing would be to sometimes pursue the option of automation, sometimes not. Indeed, there may be no-go zones for AI, not because the solutions do not yet exist, but because we value something else. Only with a foothold firmly in the soil of deep realization of the human situation; in the earthbound knowledge of the stuff we are made of and of our perennial needs and necessities, can the horizons of AI become a deeply human- and planet-centered endeavor (cf. Arendt 1958).

The endeavors to politically steer and plan must be combined with other methods of hope, such as a focus on *the human imagination.* The future demands a role for the imagination. Hence imagination and creativity are crucial for achieving an alternative that makes a difference. Jaspers explains the pivotal role of the imagination for transformation:

> It is precisely when they explain nothing and are meaningless, by the criteria of rational consequence, causality and end that myth and fairy tale can have great depth and infinite interpretability. […] *Only the language of imagination* – so it seems – touches reality that evades all objective investigation. (1937/1995, p. 83, italics in original)

Zylinska proposes, in addition, that: "[t]his possibility of envisaging a different future and painting a different picture of the world may require us to extend an invitation to nonhuman others to join the project and help redraft its aesthetic boundaries" (2020, p. 148). In order to embrace such alien epistemologies we may – in addition to turning to the 'other-than-human' realm – also embark into the neglected and alien depths of the terrains of *Existenz*. The limit situation is the long-lost relative who should be reunited with the family of human imagination, play, creativity, and aesthetic sensibility. In fact, embracing the imaginary as part of our existential practices, means to invoke the radical *openness* of the limit situation and thus to simultaneously move *beyond even that which we can imagine* (Berardi 2017). Here, the limit situation offers up a possibility to capture a neglected potentiality of being human, an alterity within our register. Hence, the alternative often sought in animals and machines, is an ultimate otherness that can also be found at the heart of what matters to us, and in our very acts of rebelliously imagining and carefully attending. Acts that evade objectivism and that may allow for a creative broadening of both the human register and our anticipatory modes and media, beyond the instrumental, logical, controlled, autonomous, certain – and in effect predicted and absolutely predictable – idea of the Human, and His Future with AI in 'the New Era'.

7 CONCLUSION

This essay set out by discussing how AI succeeds in presenting itself as that earth-shattering arrival on the human horizon at the end times, reflecting a temporal hybrid of the next and the infinite in which some forms of religious and macroeconomic discourse share a stake. This, as Shoshana Zuboff has demonstrated, includes a looting of the depths of human experience to envelope humanity's existential concerns for profit. In addition, AI entrepreneurs are in the time of writing aiming to benefit from the non-surveyable and as some would argue, interlinked crises of our present age, attempting to fill also that empty, uncertain future of the next moment with 'inevitable' datafication. One could even argue that AI imaginaries are rummaging the brinks of a destructive form of life that they simultaneously reproduce; an economic and political order that according to Adam and Groves "encourages us to fly blindly forward into the future, trusting in the protection of forecast and scientific prediction" (2011, p. 18). In other words, in an era of multiple crises, AI imaginaries – contrary to what they proclaim – continue the routine to effectively institutionalize irresponsibility, as they are "exploiting the future in the narrow interests of the present" (ibid). Presenting themselves as the only set of solutions to problems that face us on the fringes of our late modern societal order of disintegration – while operating through forecasting, prediction and precision – they thus effectively close the very horizon of the future at the same time.

An important objective has thus also been to offer an invitation beyond the prospects and limits of 'the new AI Era' of predictive modelling, exploitation and dataism. The invitation goes: let's collaboratively imagine and craft a future of existentially sustainable media. Let's pause in the present to reflect on and thus engage the future, and indeed zealously *philosophize* in the spirit of Jaspers in order to bring something else, something new, into being. Let's seek out methods of hope, beginning with the act of embracing the present moment – the digital limit situation – as a *task*. And let's pick up the torch from the *Futures Anthropologies Manifesto* for example and "probe, interrogate and play with futures that are plural, non-linear, cyclical, implausible and always unraveling" (2017, n.p.).

Precisely here lies the assignment ahead for pursuing a post-disciplinary, integrative and generative form of Humanities and Social Sciences as a method of hope, that engages AI researchers in a pursuit of designing for the benefit of an inclusive and open future of existential and ecological sustainability. Thus bridging 'the two cultures' means, I suggest, exploring an existential ethics in collaboration with those who engineer the systems, in the joint existential practice of imagining the future at the limits of what can be imagined. The digital limit situation means a chance of opening up the present to other possibilities (Bifo 2017 p. 232) than those visible, embedded, forecasted, or scientifically conceivable: to the indeterminate, open-ended, or to the completely unbelievable. As Bifo suggests, for example, the implausible scenario of a worldwide politico-ethical awakening of all the cognitive workers of the world: designers, programmers, AI engineers who control the developments – that is where a new future may begin to take shape.

It seems clear that being able to anticipate Jane Guyer's 'near future' is phenomenologically required for our common life and wellbeing, and for existential sustainability in a life of and with environmental media technologies (Peters 2015). The only way to achieve it is through a combination of plans, policies, imaginings, dreams and practices of care in the present. Thus, we need a blend of particular abstractions and carefully crafted concrete and lived futures, with AI at our voluntary disposal (!). This will imply attending and tenderly tending to, and caring for, the future in the present; practically forging a common culture (a latent future) and imaginatively producing progressive plans at the same time. In the words of Jaspers, who believes artistic ciphers can be our prod:

> Only by *attending to the ciphers of being,* can one perceive this indubitable reality; it is as if in the act of attending a transformation occurs: not only into transparency, but into the ungrounded necessity that is no longer the opposite of possibility. (1937/1995 p. 83, italics in original)

Hence, the act of *attending* is key, and this is a method of hope that will open up unforeseen possibilities. I have suggested that if we read Jaspers philosophy carefully and inventively it engenders a way to think both creatively and critically about the 'life-apparatus' of AI and autonomous systems. Pitting them against the properties of existential media enables us to ask when and how they can or cannot afford anticipation proper. I have revisited his writings on the most profound human experiences of all: the limit situations of life, where insight can be gained about what makes us human in moments of utter uncertainty and contingency, and I have sought to bring them into a conversation with our contemporary technologized culture. I chose this path not only because such profundity is in fact heavily enmeshed in the digital in a variety of ways in digital existence (cf. Lagerkvist 2019). A focus on the concept of the digital limit situation may push toward reconceiving of technology in light of a multifocal sense of *limits* – in terms of brinks, thresholds, restrictions, margins – rather than endless progress. Finally, if we reconceive of media as existential, and of existential media as anticipatory, this will complement Zuboff's ultimate remedy: reclaiming will. The existential palette is broader and more nuanced. For one thing, even as we reclaim the future tense, by our will to will, we can never be sure of the upshot. Because, in fact, in all lived-in practices "multiple dynamics interact in indeterminate ways" (Guyer 2019, p. 377). Or in Jaspers' words:

> Nobody knows where man (sic!) and his thinking are going. Since existence, man and his world are not at an end, a completed philosophy is as little possible as an anticipation of the whole. We men have plans with finite ends, but something else always comes out which no one willed. (Jaspers 1935/1997, p. 48)

Thankfully. For coexisters in their historic moment, within the confines and potentials of their technologized situation, the horizon is thus ultimately still open, impredicative and as such anticipatory. Here await fundamental, abysmal, magnificent and enormous tasks for each an everyone of us (cf. Kierkegaard 1843). And for (digital) humans "the future is not just a technical and neutral space, it is shot through with affect and sensation" and it produces "awe, vertigo, excitement, disorientation" (Appadurai 2013, pp. 286-287). In our collective and diversified digital limit situation – in itself co-constituted by technologically mediated crises, offering both limitations, contingencies and possibilities – the future also deeply matters to us. And where anticipation proper musters openness and indeterminacy, existentiality will interrupt them in deep acknowledgement also of limits. In the present moment such uncertainties as well as limits in fact carry, in their inherent inconclusiveness, a hope within.

FUNDING STATEMENT AND ACKNOWLEDGMENTS

This article is part of the research project I am heading in the Department of Informatics and Media at Uppsala University (2020-2024): "BioMe: Existential Challenges and Ethical Imperatives of Biometric AI in Everyday Lifeworlds", funded by the national programme WASP-HS and the Marianne and Marcus Wallenberg Foundation. See https://www.im.uu.se/research/hub-for-digtal-existence and https://wasp-hs.org/ I would like to thank Joanna Zylinska and Simon Lindgren for insightful comments on an earlier version of this essay.

REFERENCES

Adam, B. and C. Groves(2007) *Future Matters,* Leiden: Brill.

Adam, B. and C. Groves (2011) 'Futures Tended: Care and Future Oriented Responsibility,' *Bulletin of Science, Technology & Society* 31(1), pp. 17–27, DOI: 10.1177/0270467610391237

Andrejevic, M. (2019) 'Automating Surveillance', *Surveillance & Society* 17(1–2), pp. 7–13, DOI: https://doi.org/10.24908/ss.v17i1/2.12930

Andrejevic, M. (2020) 'Data Civics: A Response to the 'Ethical Turn'', *Television and New Media,* 21(6), pp. 562-567, DOI: 10.1177/1527476420919693

Appadurai, A. (2013) *The Future as Cultural Fact,* London: Verso.

Arendt, H. (1978) *Life of the Mind,* Vol 2. New York, Harcourt Brace Jovanovich.

Arendt, H. (1958) *The Human Condition*, Chicago: University of Chicago Press.

Arendt, H. and K. Jaspers (1992) *Hannah Arendt/Karl Jaspers Correspondence, 1926-1969*, L. Kohler and H. Saner (eds.) Orlando: Harcourt Brace Jovanovich.

Berardi, F. B. (2017) *Futurability: The Age of Impotence and the Horizon of Possibility*, London: Verso.

Boström, N. (2014) *Superintelligence: Paths, Dangers, Strategies,* Oxford: Oxford University Press.

Bucher, T. (2018) *If... Then: Algorithmic Power and Politics,* Oxford: Oxford University Press.

Chun, W.H.K. (2016) *Updating to Remain the Same: Habitual New Media,* Boston: MIT Press.

Corpus Ong. J. and D. Negra (2020) 'The Media (Studies) of the Pandemic Moment: Introduction to the 20th Anniversary Issue', *Television and New Media,* 21(6), pp. 555–561, DOI: 10.1177/1527476420934127

de Beauvoir, S. (1947) *Pour une morale de l'ambiguïté,* Paris: Galimard.

De Miranda, L., S. Ramamoorty and M. Rovatsos (2016) 'We, Anthrobot: Learning from Human Forms of Interaction and Esprit de Corps to Develop More Diverse Social Robotics.' In J. Seibt, M. Nørskov, S. Schack and S. Andersen (eds.) *What Social Robots Can and Should Do,* Proceedings of Robophilosophy 2016/TRANSOR, Amsterdam: IOS Press, 48-59.

Dencik, L. (2018) 'Surveillance Realism and the Politics of Imagination: Is There No Alternative?' *Krisis: Journal for Contemporary Philosophy* 1:31–43, https://krisis.eu/surveillance-realism-and-the-politics-of-imagination-is-there-no-alternative/

Dencik, L. (2020) 'Mobilizing Media Studies in an Age of Datafication', *Television & New Media* 21(6), pp. 568–573, DOI: 10.1177/1527476420918848

Esposito, E. (2018) 'Future and Uncertainty in the Digital Society', *Making Sense of the Digital Society,* Alexander von Humboldt Institut für Internet und Gesellschaft, 12 March 2018.

Eubanks, V. (2018) *Automating Inequality: How High-tech Tools Profile, Police, and Punish the Poor,* New York: St. Martin s Press.

Gasper, D. (2018) 'Insouciance, Indifference and Any Inspiration in the Face of Emergent Global Crises?' in B. Jessop and K. Knio (eds.) *The Pedagogy of Economic, Political and Social Crises: Dynamics, Construals and Lessons,* London: Routledge.

Gates, K. (2011) *Our Biometric Future,* New York: New York University Press.

Groves, C. (forthcoming) 'Flourishing for the Future: Anticipation as Meta-capability', in E. Spiers, M. Büscher, C. Lopez-Galviz, and A. Nordin (eds.) *Handbook of Social Futures* (title TBC).

Guyer, J. (2019) 'Anthropology and the Near-Future Concept' in R. Poli Ed. *Handbook of Anticipation,* Cham: Springer International Publishing.

Guyer, J. I. (2016) ''On the verge': From the possible to the emergent', *HAU Journal,* 6(1), pp. 373–377, DOI: http://dx.doi.org/10.14318/hau6.1.020

Guyer, J. (2007) 'Prophecy and the Near Future: Thoughts on Macroeconomic, Evangelical and Punctuated Time', *American Ethnologist,* 34(3), pp. 409-421, DOI: 10.1525/ae.2007.34.3.409

Guzman, A. (2019) 'Beyond Extraordinary: Theorizing Artificial Intelligence and the Self in Daily Life' in Z. Papacharissi Ed. *A Networked Self and Human Augmentics, Artificial Intelligence, Sentience.* New York: Routledge.

Halpern, O. (2018) "Golden Futures", *limn,* issue 10, April. URL: https://limn.it/articles/golden-futures/

Heidegger, M. (1927/1962) *Being and Time,* London: SCM Press.

Henderson, L. (2020) 'Media Studies Futures: Whiteness, Indigeneity, Multi-modality, and a Politics of Possibility', *Television and New Media,* 21(6), pp. 581–589, DOI: 10.1177/1527476420921515

Hirsch, M. (2016) 'Vulnerable Time', in J. Butler, Z. Gambetti and L. Sabsay (eds.) *Vulnerability in Resistance,* Durham: Duke University Press.

Hong, S-H. and P. Szpunar(2019) 'The Futures of Anticipatory Reason: Contingency and Speculation in the Sting Operation', *Security Dialogue,* 50(4), pp. 314–330, DOI: 10.1177/0967010619850332

Jaspers, K. (1931/2010) *Man in the Modern Age,* London: Routledge.

Jaspers, K. (1932/1969) *Philosophy, vol. I.,* Chicago: University of Chicago Press.

Jaspers, K. (1932/1970) *Philosophy, vol. II.,* Chicago: University of Chicago Press.

Jaspers, K. (1949) *The Perennial Scope of Philosophy,* New York: Philosophical Library.

Jaspers. K. (1937/1995) *The Philosophy of Existence,* Philadelphia: The University of Pennsylvania Press.

Josephides, L. (2014) 'Imagining the Future: An Existential and Practical Activity', in W. Rollason Ed. *Pacific Futures, Projects, Politics and Interests,* Oxford: Berg.

Kavedzija, I. (2016) 'Introduction: Reorienting Hopes', *Contemporary Japan,* 28(1), pp. 1-11, DOI: https://doi.org/10.1515/cj-2016-0001

Kelly, K. (2016) *The Inevitable: Understanding the 12 Technological Forces that Will Shape our Future,* New York: Viking Press.

Kember, S. and J. Zylinska (2011) *Life after New Media: Mediation as a Vital Process;* Cambridge, MA: MIT Press.

Kennedy, H. and R. Hill (2018) 'The Feeling of Numbers: Emotions in Everyday Engagements with Data and Their Visualisation', *Sociology,* 52(4), pp. 830–848, DOI: 10.1177/0038038516674675

Kierkegaard, S. (1846) 'Two Ages: The Age of Revolution and the Present Age – A Literary Review' (30 March 1846).

Kierkegaard, S. (1843) *Enten-eller. Et livsfragment,* København.

Klein, N. (2020) 'How Big Tech Plans to Profit from the Pandemic,' *The Guardian*, May 13, 2020.

Lagerkvist, A. Ed. (2019) *Digital Existence: Ontology, Ethics and Transcendence in Digital Culture*, London: Routledge. (Routledge Studies in Religion and Digital Culture).

Lagerkvist, A. (2016) 'Existential Media: Toward a Theorization of Digital Thrownness', *New Media & Society.* Online First, June 7, 2016, https://doi.org/10.1177/1461444816649921

Lagerkvist, A. (2018) 'Numerical Being and Non-being: Probing the Ethos of Quantification in Bereavement Online', in Z. Papacharissi Ed. *A Networked Self and Birth, Life, Death,* New York: Routledge.

Lindgren, S. (2020) *Data Theory: Interpretive Sociology and Computational Methods*, Oxford: Polity.

Lupton, D. (2016) *The Quantified Self: A Sociology of Self-Tracking*, London: Polity.

MacKenzie D. and J. Wajcman (1999) *The Social Shaping of Technology*, Buckingham and Philadelphia: Open University Press.

MacKenzie, C., W. Rogers and S. Dodds (eds.) (2014) *Vulnerability: New Essays in Ethics and Feminist Philosophy*, Oxford: OUP.

McQuillan, D. (2019) "AI Realism and Structural Alternatives." Paper given at Data Justice Lab Workshop, Cardiff University. http://danmcquillan.io/ai_realism.html.

Miller, R. (2007) 'Futures Literacy: A Hybrid Strategic Scenario Method,' *Futures*, 39(4), pp. 341-362, DOI: 10.1016/j.futures.2006.12.001

Mittelstadt, B. D., P. Allo, P., M. Taddeo, S. Wachter and L. Floridi (2016) 'The Ethics of Algorithms: Mapping the Debate', *Big Data & Society,* July-December 2016, pp. 1-21, DOI: 10.1177/2053951716679679

Miyazaki, H. (2004) *The Method of Hope: Anthropology, Philosophy and Fijian Knowledge,* Stanford: Stanford University Press.

Noble, S. U. (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*, New York: NYU Press.

Pentzold, C., A. Kaun, and C. Lohmeier (eds.) (2020) 'Imagining and Instituting Future Media: Introduction to the Special Issue' (Back to the Future: Telling and Taming Anticipatory Media Visions and Technologies), *Convergence,* DOI: 10.1177/1354856520938584

Peters, J. D. (2015) *The Marvelous Clouds: Toward a Philosophy of Elemental Media,* Chicago: Chicago University Press.

Pink, S. and V. Fors (2017) 'Being in a Mediated World: Self-tracking and the Mind-body-environment', *cultural geographies*, 24(3), pp. 375-388, DOI: https://doi.org/10.1177/1474474016684127

Pink, S., S. Sumartojo, D. Lupton and C. Heyes La Bond (2017) 'Mundane Data: The Routines, Contingencies and Accomplishments of Digital Living', *Big Data & Society*, 4(1), pp. 1-12, DOI: https://doi.org/10.1177/2053951717700924

Poli, R. (2017) *Introduction to Anticipation Studies*, Cham: Springer International Publishers.

Rovatsos, M. (2019) 'Anticipatory Artificial Intelligence', in R. Poli Ed. *Handbook of Anticipation,* London: Springer International Publishing.

Salazin, J.F., S. Pink, A. Irving and J. Sjöberg (eds.) (2017) *Anthropologies and Futures: Researching Emerging and Uncertain Worlds,* London: Bloomsbury.

Sartre, J-P. (1943) *L'Être et le néant : Essai d'ontologie phénoménologique,* Paris: Éditions Gallimard.

Schutz, A. (1972) *The Phenomenology of the Social World*, London: Heinemann Educational.

Sneath, D., M. Holbraad and M. Pedersen (2009) 'Technologies of the Imagination: An Introduction', *Ethnos,* 74(1), pp. 5-30.

Thornhill, C. (2002) *Karl Jaspers: Politics and Metaphysics,* London: Routledge.

van Dijck, J. (2014) 'Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology', *Surveillance & Society*, 12(2), pp. 197-208.

Withy, K. (2011) 'Situation and Limitation: Making Sense of Heidegger on Thrownness', *European Journal of Philosophy,* 22(1), pp. 61–81.

Zuboff, S. (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the new Frontiers of Power*, London: Profile Books.

Zylinska, J. (2018) *The End of Man: A Feminist Counter Apocalypse,* Forerunners. Ideas First.

Zylinska, J. (2020) *AI Art: Machine Visions and Warped Dreams*. London: Open Humanities Press.

JOURNAL OF DIGITAL SOCIAL RESEARCH

WWW.JDSR.IO

ISSN: 2003-1998

VOL. 2, NO. 3, 2020, 42–64

COPRODUCTION, ETHICS AND ARTIFICIAL INTELLIGENCE: A PERSPECTIVE FROM CULTURAL ANTHROPOLOGY

Leah Govia*

ABSTRACT

Over the past five years, artificial intelligence (AI) has been endorsed as the technical underpinning of innovation. Sensationalist representations of AI have also been accompanied by assumptions of technological determinism that distract from the ordinary, sometimes unassuming consequences of interaction with its systems and processes. Drawing on scholarship from cultural anthropology, along with science and technology studies (STS), this paper examines coproduction in a Canadian AI research and development context. Through interview responses and field observations it presents sites of sociotechnical entanglement and ethical discussion to highlight potential spaces of mediation for anthropological practice. Emerging themes from the experiences of AI specialists include the negotiability of technology, an ethics of the everyday and critical collaboration. Together this returns to an initial approach into a situated understanding of artificial intelligence, negotiating with broad, sensationalist perspectives and the more commonplace, backgrounded cases of narrow research.

Keywords: cultural anthropology; artificial intelligence; ethics; coproduction

* University of Waterloo, Canada.

1 INTRODUCTION

AI research and development has seen substantial investment in Canada. Previous government commitment has sought to position the country as "a world-leading destination for companies seeking to invest in AI and innovation"[1] and in 2017, the federal government implemented a "$125 million Pan-Canadian Artificial Intelligence Strategy, the world's first national AI strategy".[2] This, along with membership in the Global Partnership on Artificial Intelligence[3] and "fast-track visa programs"[4] for tech talent have seen multiple Canadian cities placed among the fastest growing tech markets in North America. More recently too, in response to COVID-19 we've seen the development of AI-supported contact-tracing apps contributed by companies, universities and national AI institutes.[5] It is in and among these many venues that the Canadian AI context is both flourishing locally and displaying significance internationally, at least in view of globalizing, capitalist development discourse. Acting within and between these strategies are researchers and developers whose work becomes hyper-publicized as intelligent technologies continue to enter into our daily lives. Questions about research and design further emerge in this public view, urging specialists to face the social or ethical implications of the work that they do.

While it may not be possible to ensure non-harmful application of artificial intelligence, it is important to guarantee less harmful processes in its research and development. For instance, a commonly expressed concern is the need for AI to be designed with transparency. In many domains where it integrates tangibly with varying publics and stakeholders, such as in the health and financial industries, transparency has become synonymous with trust and accountability (Kim et al. 2014; Manderson et al. 2015). When seeking such transparency, it is necessary to understand how specialists engage with dynamic, sociotechnical articulations — in and of their work — as a nexus where ideas of trust and accountability also configure and emerge. Here too, as AI is drawn into common social, political, public venues, anthropological mediation becomes useful when accounting for negotiation between the imagined and realized sociotechnical contexts specialists grapple with.

[1] Government of Canada https://www.canada.ca/en/department-finance/news/2017/03/growing_canada_sadvantageinartificialintelligence.html

[2] CIFAR https://www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy

[3] Global Partnership on AI https://www.therecord.com/news/waterloo-region/2020/06/16/canada-joins-international-partnership-to-promote-responsible-ai.html

[4] Fast-track visa programs https://dmz.ryerson.ca/the-review/artificial-intelligence/

[5] TraceSCAN https://uwaterloo.ca/stories/news/new-ai-technology-will-be-used-improve-contact-tracing-covid; Mila https://globalnews.ca/news/6951846/coronavirus-contact-tracing-app/

Within anthropology, literature on artificial intelligence is still emerging and less established than in other areas of STS, but topics such as post-humanism, virtual worlds, human-machine interaction, big data and algorithms offer related insights (Born 1997; Robertson 2010; Nardi 2010; Boellstorff et al. 2012; Richardson 2015; Irani 2015; Seaver 2018). More specific to AI, an earlier inquiry by Mariella Combi (1992) focuses on the AI imaginary to display how problems and solutions, both technical and social, are constructed through human-computer relation. Similarly, the late Diana Forsythe's work during the early 1990's involves an ethnographic account of knowledge-making in an AI scientific community. She investigates shared practice and meaning to present how knowledge is localized rather than representative of a universal commonsense (Forsythe 1993ab). This is further accompanied by an extensive body of literature from science and technology studies (STS), which aims to critically examine the construction of scientific knowledge and practice. Through this field of study one can investigate the interplay between "epistemic and political processes" to demonstrate how technologies and social orders are co-produced. Including theory on the agency of things, when extended to artificial intelligence STS considers symbolic and material agencies that transform spaces, facilitate experience and create different kinds of relations — sociocultural, ethical, technical or otherwise — through coproduction (Latour and Woolgar 1986; Haraway 1988, 1990; Latour 1991, 1999; Hacking 1999; Bille and Sorensen 2007; Solomon 2008; Sismondo 2008; Ingold 2008; Jasanoff 2004, 2016). To simply illustrate, the programming decisions that specialists make when coding and the computational agencies of algorithms that arise as such, while typically viewed as technically independent are situated with certain historical, cultural or political arrangements that already inform available choices and potential outcomes. Which theories and algorithms are framed as most suitable for varying software applications, how data is represented, or the ways in which coding practices come to "matter"; these emerge through the interrelation of technical practice and the particular contexts where specialists construct knowledge of said practice (Reardon 2001).

An anthropological perspective is suited for sociotechnical analysis or for identifying sites of coproduction. It provides reflexive understanding that phenomena are all at once situated, dynamic, emergent, and in this seemingly conflicted, yet grounded plasticity is the presence of negotiation. When accessed in the creation of regulatory frameworks and policies, for example, it foregrounds a heterogeneity of publics and stakeholders for intelligent technologies constructed to suit a wide range of experiences and contexts, not only those that reproduce hegemonic, normative subscriptions of being (Mosemghvdlishvili and Jansz 2013). An anthropological approach concerned with situated knowledges and embedded action may help to manage the technological landscapes that influence how we interact with our worlds, sometimes in ways we've yet to imagine (Haraway 1988).

2 METHODS

Data collection was supported by methods including semi-structured interviews, unobtrusive observation, archival research and textual analysis. Semi-structured interviews were conducted both in-person and virtually, guided by questions that asked respondents to share their experiences working within the field of artificial intelligence, or working with any of its associated techniques for cognate disciplines such as quantum computing. Respondents were also asked to discuss the social or ethical implications of artificial intelligence, both particular to their work and to more publicized examples. This ranged from industry professionals using machine learning techniques for business strategy, to graduate students exploring ethical algorithms in their dissertation. Interview audio was then transcribed and thematically coded, manually and with analysis software Atlas.ti.

An academic conference, AI guest talks and group meetings were the primary sites of unobtrusive observation. For example, at the Fifth Annual Conference on Governance of Emerging Technologies: Law, Policy and Ethics, I attended talks presented by Canadian and international researchers on topics specific to regulation, ethics and artificial intelligence. Entering these spaces and "becoming the phenomenon" or attempting to simulate a position similar to that of the specialists attending increased access to epistemological processes that membership within an AI-related community might afford (Franklin and Roberts 2006). Standard to an anthropological approach, this interpretive process is informed by a notion of ethnography as embodied practice and highlights the dynamic activity of the field (Cerwonka and Malkki 2008). Key respondents were later identified and include computer science professors, PhD students/candidates, a post-doctoral researcher and an industry professional (P1, P2…P7).

While each individual worked within an AI community or related space, particular emphasis was placed on those based at a Canadian institution in Southwestern Ontario, Canada. The faculty of computer science at this university is renowned for its connections to the tech industry, with graduates often finding placement in positions at companies such as Apple, Facebook, and Google. Existence of an Artificial Intelligence Group and its most recent collaboration with the Partnership on AI further evidences a concentration of AI research at the university, adding to its appeal as a source of data. Interviews were approached with the concept of engaged listening and an ethnographic imaginary meant to provide insight similar to that of participant observation (Forsey 2010). For additional data, basic textual analysis was applied to public policy recommendations, reports, and design guides from two North American R&D organizations with designated initiatives that address the social implications of artificial intelligence. These are the Canadian Institute for Advanced Research (CIFAR) [6] and the Institute of Electrical and Electronics Engineering Standards Association (IEEE SA)[7].

3 UNDERSTANDING COPRODUCTION AND AI

Artificial intelligence has been categorized in different ways to distinguish function and capability, with terms such as "weak"/ "narrow" and "strong" AI being used, although these categories overlap and are not consistently taken up by researchers (Warwick 2013). The specialists I spoke with predominantly worked on narrow AI, which has been described as deliberately programmed, task-specific, or with capabilities restricted to a single domain (Bostrom and Yudkowsky 2014). When discussing their thoughts on the discipline, a common theme among my interlocutors was that artificial intelligence is nowhere near the level of capability displayed in the media. As one PhD student succinctly explained: "public conversation glosses over critical distinctions in what's actually possible, and what we foresee as feasible, and what's currently being done" (P3). Though others confirmed that various forms of artificial intelligence can and will continue to surpass human performance, as intended, they also echoed the words of the PhD student with reference to current applications of AI being single-purpose. One doctoral candidate recounted their conversation with a "bleeding edge researcher", stating that from a few years ago:

> The bleeding edge development is the robot can figure out when a chair is in its way, and move the chair out of its way so it can continue rolling down the hallway…so if AI were to take over the world you would not be able to stop them by putting chairs in their way…not that particular model of chair. (P2)

These specialists are aware of the sensationalist expectations crafted with public understandings of AI, in coexisting, historicized and emerging sociotechnical imaginaries, but it may not align with the technical realities of their work. The non-technical is sometimes placed external to these realities too. Here I return to the foundational suggestion that distinctions between the social and technical are often fabricated rather than actual. Technology is not constructed in isolation, but instead co-produced with

[6] CIFAR https://www.cifar.ca/ai

[7] IEEE SA https://standards.ieee.org/industry-connections/ec/autonomous-systems.html

"social practices, identities, norms, conventions, discourses, instruments and institutions" (Jasanoff, 2004, p.3; Latour and Woolgar 1986). Identifying sites of coproduction in artificial intelligence can expose how its features are in constant entanglement while simultaneously emphasizing said features and the ways they hold potential, contingent configurations specific to the AI context, while moving beyond narratives of technological determinism.

Studies in educational settings show that like other subfields of computer science, AI is considerably practice-oriented (Kay et al. 2000). While it is seemingly obvious to state, students learn various coding languages and become familiar with how developer input influences the functions of a system. With primary actions virtually facilitated through a computer, to specialists "the central meaning of work may be writing code and building systems" (Forsythe 1993a, p. 470). This was similarly noted by a professor of computer science who explained that in AI, "at the research level it's just studying algorithms", sharing an example where "you create some image database and then you write some algorithms to classify images or something like that, but you can do all that without asking a human being to do anything" (P5). There are moments in daily practice that are conventional and distance specialists from the sociality of their work, but when the characterization of work in AI is assigned to certain structures of discourse, other topics can be sidelined or positioned as external to the technical aspects in focus (Forsythe 1993ab). It may also mask other considerations and consequences of the technologies at work:

> Artificial intelligence has always been concerned primarily with building machines that are operating independently from humans. Most of AI is building machines that have nothing to do with human beings, they're just completely separate. Even a machine that plays chess, it doesn't care that it's playing against a human. It could be playing against another machine; it's got no model of the human. Same thing for these poker-playing robots. They're not modelling human feeling they're not modelling human anything; they're just modelling the game. They're just modelling inanimate objects and that's all…that's really weird when you think about it. There's no doubt that everybody must know that intelligence has a lot to do with other people. (P5)

As the quote above indicates, the professor is attuned to a real and imagined social presence of artificial intelligence, but the positioning of AI as a technical object is, as appropriate to the discipline, most attended to. This removal of the "human variable" is a more pronounced display of how the separation of social and technical aims to leverage the universality and "effectivity" of technology (Born 1997). At the same time, this universality provides space for technologies to be aligned with larger structural and institutional goals, which contradicts the supposed separation that sources its universality. Such a view is not uncommon and follows the positioning of science external to the social to "protect the 'value neutrality' of the scientific process" (Douglas 2007, p.127; Liu 2017). While many of the practical, operational aspects of AI appear to be separated from humans with an emphasis on features like automation, for instance, the development of automation has always been fundamentally entangled and co-productive. Even in systematic categorizations of autonomy considering independent, agential action separate from a programmer's original input, there remain many scenarios in which developers must evaluate and re-adjust the machine's operational capacities (Warwick 2013; Richardson 2015). It is because technology is shaped by constraints or conditions in design and application that technical decisions made at one point in time can impact development made at another, or vice versa (Mosemghvdlishvili and Jansz 2013). This reconfirms that in the pathways of research and development, from acquiring datasets and programming algorithms, to designing user interfaces and eventual implementation, AI is in constant coproduction.

3.1 Making the social, technical

At the conference on Governance of Emerging Technologies where part of my observation took place, during a keynote speech the founder of the Center for Human-Compatible Artificial Intelligence [8] called to both "maximize human values" and manage risk in AI by accounting for the "biggest deviation of rationality" — our wants. He expressed that by learning to predict what people want, it will become easier to develop systems that are beneficial and will require "cultural work" to reach prediction. The call for cultural work seemed to suggest a holistic survey of societies globally for a shared set of wants, following research on the use of psychological and sociological modelling for artificial intelligence. For example, in the subfield of Affective Computing a major theoretical influence comes from Affect Control Theory (ACT). This sociological theory considers the relationship between emotion and culture, categorizing patterns of affective meaning that are socially shared (Rogers et. al 2014). One of my respondents is a professor who works with building such sociological models into AI solutions through this field of research. They explained that the modelling relies on "the sort of collective consciousness or collective nature of human intelligence", which in this case is associated with affect and emotion (P5). This is then mapped to cultural contexts through AI techniques. One such mapping is exemplified by a program the

[8] Center for Human-Compatible Artificial Intelligence https://humancompatible.ai/

professor has drawn influence from in his research, known as Interact. Available for download through Indiana University (2016):

> Interact is a computer program that describes what people might do in a given situation, how they might respond emotionally to events, and how they might attribute qualities or new identities to themselves and other interactants in order to account for unexpected happenings. Interact achieves its results by employing multivariate non-linear equations that describe how events create impressions, by implementing a cybernetic model that represents people as maintaining cultural meanings through their actions and interpretations, and by incorporating repositories of cultural meanings.

The repositories of cultural meanings are formatted as dictionaries of affective meaning. These contain set identities, behaviours, and settings. Categorized by place and date, some of the listed dictionaries include U.S.A.: Indiana 2003, Japan 1989-2002, Germany 2007, and Northern Ireland 1977. Data from these dictionaries then help to model interactions between actors and objects as events and determine the probable impressions each person holds after certain event actions. Cultures are depicted as totalities within Interact and fall within a normative process supported by philosophies of science that emphasize naturally embodied dispositions substantiated by a group, corroborated as "culture". Anthropologists, however, have problematized the definition of culture as a bounded concept. Emphasizing intragroup variations and movements, they argue that cultures are not homogenous entities. [9] The categorization in Interact of place-based identity, behaviour and setting meant to determine affect and impression reproduces the definition of culture as bounded and assumes a universality of emotions. It also places social experience as something that is rigidly patterned, based on its representation as static and deterministic. Instead, the codifying of emotions is already bound by cultural interpretations of emotion in the Interact program because it is influenced by the epistemological stance of ACT. In the representations of consistent "cultures", it also simultaneously erases and reifies various social and cultural elements due to a reliance on universality. Again, anthropologists have problematized universality and homogeneity both theoretically and methodologically. An added ontological viewing would further question the universality applied to social and cultural phenomena in Interact. These phenomena are brought into existence through their delineation in the first place, rather than being universally attributed, pre-existing conditions (Coopmans et al. 2014; Hoeppe 2015). In other words, the codifying of

[9] Definitions of culture have been problematized for many years (Gupta and Ferguson 1997; Hobart 2000; Clifford and Marcus 1986; Helmreich 2001)

culture and emotions in Interact is an embodied cultural interpretation — a phenomenon brought into the world through the activity of coding itself.

While this example from the professor is a plainly demonstrated site of coproduction, others are not as immediately discernible. As sociocultural factors are datafied, they become inscriptions: "visual/textual translations and extensions of scientific practice" that frame said factors as technical objects to legitimize their presence (Latour and Woolgar 1986, p.142). In making the social, technical, these essentialized, deterministic evaluations of sociocultural phenomena appear. Alternatively, going "back to the basics" in an anthropological or STS approach that calls attention to coproduction is not just a reminder, but an available strategy for interested specialists who find concern with the structuring of data or algorithm design. Within their work specialists do craft an understanding of the sociotechnical, where systems articulate with other forms of expertise and knowledge, all of which is value-laden. They balance a range of factors including technical operations, funding influences and design compatibility while ensuring that their work is adapted for other, already existing emerging technologies and the various contexts where AI is applied (Ekbia 2008; Johnson and Wetmore 2008). It is understandable that the keynote speaker mentioned a need to both maximize human values and deal with risk in AI. Exactly how our values are being handled still needs care-full, reflexive consideration and increased interdisciplinary collaboration, as many have already called for.

Importantly, it also asks us to confront the difficulties of making AI socially and technically sustainable. Programs like Interact may begin as an exploratory project in mapping moments of human sociality, but when implemented more widely, present worries similar to that seen in cases of algorithmic bias and imbalanced datasets. Concurrently, they call on the agential capacity of AI that generates a seemingly separate, yet impactful trajectory of more-than-human expansion. The sense of agency that AI evokes, especially when projecting affective qualities, is then heightened in social perception of its systems. Aligned with studies on anthropomorphism and technology, this suggests that specialists may unintentionally act to maintain their social worlds in research and development to more easily maneuver the unpredictable relation to more-than-human agencies (Eyssel et al. 2012; Picarra et al. 2016). Common exposure to such "affective algorithms" might long remain more speculative than practical, but as initially noted, seeing to sustainable sociotechnical relations at minimum requires us to acknowledge the messiness of coproduction, from conceptualization to application. Anthropologists can contribute with further analysis and ethnographic endeavors that showcase the situatedness of what it means to "do" AI, with and beyond human relationality, while offering tools of accountability through critical reflection on the ontological and epistemological conditions in research and development.

4 CONSIDERING ETHICS AND REGULATION

In its most public standing, the ethics of artificial intelligence is an applied ethics. Among the focus on implications or consequences, academic inquiry has also specified a combination of theory and application through subfields such as roboethics and machine ethics (Wallach et al. 2008; Dougherty 2013; Vanderelst and Winfield 2018). In practice, discussions tend to privilege certain configurations or models of ethics, mainly those influenced by European moral philosophy that frame ethics as a complex form of decision-making (Torrance 2013; Englert et al. 2014; Cervantes et al. 2016). This was further confirmed by multiple respondents when the topic of AI ethics was raised, like a PhD student specializing in computer vision noted:

> If you're familiar with various philosophical theories of ethics, a lot of them involve either satisfying constraints based on rules, Kantian deontological ethics, or optimizing some function, Utilitarian like Mill or Bentham…now these sorts of optimization are actually very important in computer science in general, also in artificial intelligence. (P3)

One way this fits within experiences of AI ethics is through a framing of complexity and the popular narrative of innovation being inherently beneficial to humanity guiding research and development (Ekbia 2008). Both professors (P4, P5) mentioned a pattern in AI, like other STEM-related disciplines, where certain breakthroughs reach a level of visibility that sparks interest in the public. The current interest surrounds work on machine learning and deep neural networks, but they explained that this happens "once every 10 years" and that there have been at least "two of these hypes in the past" for AI. The professor whose research involves constraint programming described this as techno-optimism. They shared that the view of technology solving "all the problems and it's only a good thing" is a regular interpretation at their campus when students or colleagues discuss the social implications of artificial intelligence (P4).

This was further illustrated after a guest talk by the previous director of Microsoft Research Labs, Eric Horvitz. Speaking of the then-director's proposal about autonomous driving as the solution for deaths by drunk driving, the professor shared:

> There's easy technological fixes that prevent people from driving their cars when they're drunk…You don't have to go to autonomous driving to save 40,000 people, you can do it for a few hundred dollars. Autonomous driving will add thousands and thousands to the price of a car, so it's more of a 'I

> love technology' thing as opposed to a rational decision about what's the best way to prevent these deaths. (P4)

Here he suggests that there are already existing, commonplace fixes for current problems, but they are overshadowed by techno-optimism and to some extent, a fetishization of innovation. It is a sentiment that is similarly seen with the "black box" problem in AI, where the internal operations are mostly unknown, yet the output or outcomes — when they appear to be useful or harmless — can be left unchallenged. Though the black box problem is exacerbated by an amount of data and processing too complex for individual understanding, the complexity it engenders also motivates a simpler viewing of technology as isolated. Data is usually highlighted as the likely vessel for bias in this scenario, given its more direct connection to developer input and decision-making, but algorithms are no more separate from their sociotechnical makings than the data that feeds them. With algorithms learning from "either the human-trained input or the self-learned input" specialists aim to "identify what those outcomes are of the algorithm" (P6). But as previously shared, these outcomes can mirror social orders by the very act of their structuring whereby certain technical solutions become entrenched with choices determined and made available to groups with specific social, political or economic power. Combined, this has already translated to outcomes in facial recognition technology and predictive policing that reproduce existing inequalities, largely expanding on colonial makings that continue to place Black, Indigenous and racialized communities under directed surveillance by the state (Buolamwini and Gebru 2018; Benjamin 2019). Along this view, the black box of techno-optimism where technological success masks the intricacies of research and development prompts specialists to focus on those same tasks that create concern in the first place.

Still, to many of my respondents, these tasks do not intentionally fall into the black box of techno-optimism, they merely follow what it means to do work in artificial intelligence. Here between the hype things seem a bit more mundane, but are an important point of entry for discussions on ethics. Referring to students in the undergraduate courses he teaches, a doctoral candidate explained why such discussion is sometimes hard to find:

> Their jobs are not going to be 'how to design a comprehensive framework for running autonomous cars as a company, as a societal thing'; it's going to be 'can we solve this route planning problem for autonomous cars? Can we do image recognition accurately? And these are extremely important pieces of the puzzle, but it's not the part of the puzzle that touches on ethics. And so getting them interested in it would be difficult. (P2)

This does not mean that specialists have a lack of interest in ethics or ethical discussion. It instead confirms that work in AI is characterized according to structures of discourse that traditionally emphasize the technical prominence of the field, as was examined earlier. Again, this returns to the usefulness of identifying coproduction, particularly as the doctoral candidate's example introduces how ethics is positioned as something external to the technical. Both faculty and graduate students similarly suggested that ethical discussion is considered a "challenge outside of the curriculum", or is done in an "intentional way" through workshops outside of their research (P1). One of the professors additionally suggested that it may be because "people don't like to look too closely at what they're doing I guess, 'cause it's troubling sometimes, the role that we play" (P4). For some this translates into a question of understanding or competency:

> It's hard for me to talk about ethics because I don't really understand it that well to be quite honest with you; and that's probably the same for a lot of computer scientists, artificial intelligence researchers — that we're not too clear on what ethics is. I'm trying to learn, understanding it now at this kind of cultural consensus about things that we label as good vs bad essentially, but I know that there's other aspects to it. There's these 'whether you believe that all that matters are the consequences of things', what are these deontological ethics or consequential ethics. (P5)

> I don't know if I am qualified yet to really make professional thoughts about it. I don't have an ethics background. I have a computer science background which maybe gives me insight into some areas of it, but certainly does not give me the full picture. (P2)

In the above, ethics is discussed according to some form of formalized model of thought, either as philosophical theory, or as a professional background in ethics. There also exists a designation of authority for whom may discuss ethics and how it should be done that aligns with ethics as a delegated field of study. This is further supplemented by an underlying theme of uncertainty. For these AI specialists, uncertainty can be framed as both a challenge within the technical side of computer science and based on their responses, one that is ethical. On the technical side, there is the problem of "reasoning under uncertainty" that is and "has always been a key challenge in artificial intelligence" (P3). The other challenge is uncertainty that accompanies the ethical dimensions of emerging technologies and becomes normalized through the placement of ethics as external to practice, or as an add-on that specialists are not positioned to access (Akama et al. 2015). The dominant presence of ethics as an independent field of expertise and a major source of uncertainty, when taken up by specialists facilitates a detachment from ethical practice despite being deeply implicated in ethics-focused structures of discourse, sensationalized or otherwise.

As this exploration of ethical practice comes with a partially normative approach, it is helpful to address a context where ethics enters narrow AI research and development more explicitly. In the subfield of machine ethics (ME) there is focus on ethical embodiment by intelligent machines (Brundage 2014; Vanderelst and Winfield 2018). Value judgments of morality are referenced here and often follow the two theories of ethics my respondents mentioned: deontological and teleological. While these models form top-down and bottom-up approaches, because of the functions they feature (e.g. utility function), there are issues with constraints and optimization where "some rather technical properties of the function" make it "very hard to find the best solution" (P3). Trade-offs between different group interests are one such complication depending on the functions used (P3). Also interesting to note, is that machine ethics recognizes the agency of AI and the extent to which it catalyzes ethical practice. AI agents are categorized here as implicit or explicit to indicate the "source" of ethics, either from the designer or from the machine's self-learning (Anderson and Anderson 2007; Veruggio and Abney 2011). Thus, in one way ME queries the performance of human ethics acting upon machine and in another it holds concern for AI as an independent, ethical agent. From both, it is as if human and machine intertwine through a sociotechnical ethic where the very relation to another entity designates an "implicit moral relationship" (Scheper-Hughes 1995). Extended to the broader discussion on ethics, this reintroduces questions of accountability when facing harmful AI outcomes and forwards action for a new set of sociotechnical, legal precedents, rights, and debates on the positionality of technologies by those deemed responsible for the public good (Ekbia 2008; Nota 2015).

In any event, whether for ethical AI or an ethics of AI, it is possible to tend to separations of ethical practice and recall coproduction by highlighting some of the ways that AI specialists configure the ethical in the everyday. As an anthropology of ethics this seeks to understand how morality is manifested and maintained in the range of experiences, contexts and interactions of individuals, agents and communities, for themselves and with others (Zigon 2010; Lambek 2010). It emphasizes how morality is not a closed system, but is relational and radically context-dependent. In this way, even uncertainty becomes an ethical relation. Given a gap in the literature on AI ethics and anthropology, further studies are needed to strengthen this approach, but a focus on ordinary, everyday practices and their ethical relations is one place to start. This can rely on ethnographic and participatory research in AI contexts, providing insights on how the ethical is situated in certain positionalities, where sites of coproduction emerge, and how this interlocks with features surrounding governance, public trust in science or responsible design, to name a few broad examples. Engaging in this way could create and reconfigure choices in the negotiability of AI that expand access to the research and development of artificial intelligence, demonstrating transparency and building trust.

4.1 Fitting in regulation

When talking of ethics and regulation it is difficult to introduce weak/narrow AI without the influence of strong AI. Currently, the former is more practically delimited in discourse because application outcomes reach foundational structures and social institutions such as labour, security and healthcare.[10] The latter frequents sensationalized displays given its historicized presence in science fiction throughout literature and popular media, but still feeds into public communication of AI development, weak or strong. Together they foster anxious imaginaries in the public and include a perceived loss of agency where automation techniques are continually "becoming on-par or even better than human experts" (P6). Although P6 is referring to efficiency or accuracy in technical tasks, this "better than human" notion may in fact shift AI into positions of increased authority and affect how we orient ourselves with the world around us (Turner 2007; Muller 2014). Of course, among such sociotechnical barriers are opportunities to unsettle the conditions that arrange social, technical, or even ecological phenomena within hierarchies of value. Through the avoidance of "both social and scientific determinism", once more STS and anthropology supply space for and attention to alternative forms of regulation that might be viewed as a means of reconciling the many types of agency and artificial intelligence (Irwin 2008).

Frameworks and standards for ethical practice guiding narrow AI research and development are varied and localized. In speaking with respondents and from my field observations there are informal forms of best practice or documents employed through their local affiliations, but anything encompassing does not appear to be feasible. Here, themes of restraint and accountability emerge. A co-founder and CEO of an AI start-up in the talent acquisition industry preferred the term moderation rather than regulation, explaining that it would be better not to stop the "trajectory of technology" this way (P6). Techno-optimism appears again in his response, along with notions of technological determinism and isolation that were examined in previous sections. Simultaneously, specialists navigate the ethical urgency that emerges with AI. This comes out in conversations on regulation that are worn with uncertainty, especially at intersections of ethics and governance. Certain topics reasonably dominate

[10] Telehealth and artificial intelligence https://techcrunch.com/2017/04/19/ada-health/

due to their high-risk characterizations, like lethal autonomous weapons and technological unemployment:

> For some reason we're okay with people killing other people, but having an AI agent decide to kill a person people are less comfortable with (P2)

> In my opinion, AI is going to kill people. Not in the way that everyone thinks it's going to kill people, but people are going to die because of artificial intelligence. There is going to be job loss and it's going to be rapid and rampant. Now, the whole idea of people saying, well 're-skilling, re-training' that means the upending of an entire ecosystem called our current public education system which hasn't been revised since the first industrial revolution when it was generated (P6)

To implement regulation in these areas, many specialists outlined the importance of collaboration between politicians, legal scholars, application area experts and other AI specialists when creating frameworks or policies. A post-doctoral researcher interested in quantum computing and neural networks also showed how this compares with their regular interactions in the research community:

> It's also a very insular community right, like I personally know people at all of those companies, at high-up positions, and I'm like 3rd year post-doc I'm not a super senior person. My bosses personally know the founders of the groups at those companies right, so it's a very close-knit community that everybody knows everybody. So you're almost self-regulating just by the fact that the community is so small (P7)

This may indicate a slight disconnect between some research communities and larger planning for regulation, but as another respondent reminded "one thing that people often don't think of in the general public discourse is that somebody is going to have to actually write the programs that do these things"; that ultimately, those involved will have to listen to computer scientists about what is computationally possible (P3). It is an important consideration, but as this paper has also reminded, there are more than computational factors that will affect what is possible. Evidently, community and collaboration both influence what is possible in regulation and becomes a form of regulation itself. Collaboration may also act as ethical practice through its relationality. In North America, this has been visible in initiatives by organizations like The Canadian Institute for Advanced Research (CIFAR) and the IEEE. In the Canadian context, federally funded CIFAR is heralded for its interdisciplinary research and global network of commitments as it leads the country's national AI strategy. They produce policy recommendations and reports while creating special-interest workshops for programs analyzing AI and society. The American-based IEEE also has defined output through its Global Initiative on Ethics of Autonomous and Intelligent Systems. Both strongly advocate a collaborative approach to ethics and regulation, emphasizing thought leadership across academia and industry.

Through basic textual analysis of Building an AI World[11], Rebooting Regulation[12], and Ethically Aligned Design[13], there appears to be limited inclusion of thought leaders in groups beyond dominant technical, legal, psychological and economic backgrounds. Little reference is given to contributions by historically underrepresented communities, researchers and practitioners who have already evoked many of the themes explored in this paper involving sociotechnical coproduction, situatedness and critical reflexivity (Gasparotto 2016; Winchester III 2019; Mohamed et al. 2020). It is here that the significant underrepresentation of Black, Indigenous, and racialized persons, women in particular, is once again apparent. Underrepresentation in STEM has been well-documented and despite equity initiatives continues to persist while certain ontological and epistemological conditions tolerate a critical lack of reflexivity (Morganson et al. 2010; Fontana et al. 2013). Certainly, the insular community referred to by the post-doc is not a revelation; neither is its presence being replicated in regulation. As mentioned, even in plans claiming “diversity and inclusion”, peoples, knowledges and ways of being remain excluded because source institutions are not distanced from the underlying structures or discourses that background not only their DEI initiatives, but positionalities in leadership, e.g. related histories, identities, and practices (Ahmed 2007). Similarly, in the documents noted earlier, conventional expertise and disciplinary boundaries structure access and standards for ethical discussion. It reinforces certain definitions of ethics, questions of ethical practice, and what the major social implications of AI are (Reardon 2001). Choices and solutions are similarly narrowed and limited.

By “going back to the basics” in this way with an understanding of coproduction through critical inquiry, it may become easier to avoid such black-boxed conditions for AI ethics and regulation. The post-doctoral researcher shows how it might occur, even if limited, given their recognition of how insular the research community is and the way it translates to “self-regulation”. Alternate action acknowledges space for critical collaboration, though this requires future analysis to substantiate. Finally, critical collaboration as regulatory practice includes multiple

[11] Building an AI World https://www.cifar.ca/docs/default-source/ai-society/buildinganaiworld_eng.pdf

[12] Rebooting Regulation https://www.cifar.ca/docs/default-source/ai-reports/rebooting-regulation-exploring-the-future-of-ai-policy-in-canada.pdf?sfvrsn=616c04f3_8

[13] Ethically Aligned Design https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf

publics beyond thought leadership in academia or industry. It comes back to a request across disciplines, anthropology included, for engagement with non-specialized communities beyond research participation, although institutional and funding restrictions may hinder efforts in knowledge translation (Hayden 2007). The earlier discussed efforts by CIFAR and the IEEE do acknowledge this too, but it is not yet clear how accessible their feedback processes will be. Moving forward, the notion of critical collaboration presented here is just one of many considerations that have informed or acted alongside potential regulatory practices but may need reassessment throughout AI research and development.

5 CONCLUSION

Set in a Canadian context, this paper investigates coproduction and artificial intelligence from an anthropological perspective and is supplemented by foundational STS theory. Through noticeable examples of coproduction, first I introduce an anthropological approach to sociotechnical analysis of artificial intelligence, including the negotiability of technology. Interview responses and field observations specifically highlight the experiences of AI specialists and the ways in which sociocultural and technical elements entangle in the everyday. Next, AI ethics is situated with an anthropology of ethics through discussions on techno-optimism and conditions of uncertainty. Finally, accompanied by basic textual analysis of CIFAR and IEEE documents, regulation and ethical practice are addressed with the recommendation of critical collaboration that calls for additional reflexivity in public R&D practice.

It is important to note that this research is limited, particularly by a small sample size and reduced observation timeframe. As a result, the primary data can only represent a specific, localized Canadian context aligned with those already interested in the present topic. Despite such limitations, it acts as an initial return to a situated understanding of artificial intelligence and proposes further analysis from anthropological perspectives. It also indicates how STS can help to navigate the tensions that emerge when technical decisions are at odds with their wider social contexts. This is most noticeable in public perceptions of AI where imagined possibilities are complicated with the realities of technology. Again, additional study is surely required to go beyond my brief focus on a small grouping of specialist experiences in artificial intelligence, to the great variety of communities, discourses and processes that continue to emerge. It would be encouraging to see future works include ethnographies of applied AI and public knowledge settings, feminist analysis of AI systems in healthcare, or perhaps participatory action research on globalizing AI governance. A digital ethnographic study of machine ethics, the field focusing on ethical embodiment by intelligent machines, might also be of interest (Anderson and Anderson 2007). In our attempts to secure both equitable and non-harmful outcomes from artificial intelligence, returning to a basic, but critical understanding of sociotechnical coproduction, along with how we reach this understanding is important.

FUNDING STATEMENT AND ACKNOWLEDGMENTS

This paper draws on research supported by the Social Sciences and Humanities Research Council (Canada).

REFERENCES

Ahmed, S. (2007). ”You end up doing the document rather than doing the doing”: Diversity, race equality and the politics of documentation. Ethnic & Racial Studies, 30(4), pp. 590-609. https://doi.org/10.1080/01419870701356015

Anderson, M., & Anderson, S. L. (2007). Machine ethics: Creating an ethical intelligent agent. AI Magazine, 28(4), pp. 15-15. ttps://doi.org/10.1609/aimag.v28i4.2065

Akama, Y., Pink, S., & Fergusson, A. (2015, April). Design+ Ethnography+ Futures: Surrendering in Uncertainty. In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, pp. 531-542.
https://doi.org/10.1145/2702613.2732499

Benjamin, R. (2019). Race After Technology: Abolitionist Tools for the New Jim Code (1st ed.). Polity.

Bille, M., & Sørensen, T. F. (2007). An anthropology of luminosity: The agency of light. Journal of Material Culture, 12(3), pp. 263-284. https://doi.org/10.1177/1359183507081894

Boellstorff, T., Nardi, B., Pearce, C., & Taylor, T. L. (2012). Ethnography and virtual worlds: A handbook of method. Princeton University Press.

Born, G. (1997). Computer software as a medium: Textuality, orality and sociality in an artificial intelligence research culture. In Rethinking Visual Anthropology, pp. 139-69.

Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. In The Cambridge Handbook of Artificial Intelligence, 1, pp. 316-334.

Brundage, Miles. (2014). "Limitations and risks of machine ethics." Journal of Experimental & Theoretical Artificial Intelligence, 26, pp. 355-372. https://doi.org/10.1080/0952813X.2014.895108

Cervantes, J. A., Rodríguez, L. F., López, S., Ramos, F., & Robles, F. (2016). Autonomous agents and ethical decision-making. Cognitive

Computation, 8(2), pp. 278-296. https://doi.org/10.1007/s12559-015-9362-8

Cerwonka, A., & Malkki, L. H. (2008). Improvising theory: Process and temporality in ethnographic fieldwork. University of Chicago Press.

Clifford, J., & Marcus, G. E. (Eds.). (1986). Writing culture: the poetics and politics of ethnography: a School of American Research advanced seminar. Univ of California Press.

Combi, M. (1992). The imaginary, the computer, artificial intelligence: A cultural anthropological approach. AI & Society, 6(1), pp. 41-49. https://doi.org/10.1007/BF02472768

Coopmans, C., Vertesi, J., Lynch, M. E., & Woolgar, S. (Eds.). (2014). Representation in scientific practice revisited. MIT Press.

Dougherty, M. (2013). Something Old, Something New, Something Borrowed, Something Blue Part 2: From Frankenstein to Battlefield Drones; A Perspective on Machine Ethics. Journal of Intelligent Systems, 22(1), pp. 1-7. https://doi.org/10.1515/jisys-2013-001

Douglas, H. (2007). Rejecting the Ideal of Value-Free Science. In Value-Free Science? Ideals and Illusions. Oxford: Oxford University Press.

Ekbia, H. R. (2008). Artificial dreams: the quest for non-biological intelligence. Cambridge University Press.

Englert, M., Siebert, S., & Ziegler, M. (2014). Logical limitations to machine ethics with consequences to lethal autonomous weapons. arXiv arXiv:1411.2842

Eyssel, F., Kuchenbrandt, D., Hegel, F., de Ruiter, L. (2012). "Activating Elicited Agent Knowledge: How Robot and User Features Shape the Perception of Social Robots." Robot and Human Interactive Communication, pp. 851-857. doi: 10.1109/ROMAN.2012.6343858.

Fontana, M., Wells, M. A., & Scherer, M. C. (2013). A holistic approach to supporting women and girls at all stages of engineering education. Proceedings of the Canadian Engineering Education Association (CEEA). http://ojs.library.queensu.ca/index.php/PCEEA/article/view/4873

Forsey, M. G. (2010). Ethnography as participant listening. Ethnography, 11(4), pp. 558-572. https://doi.org/10.1177/1466138110372587

Forsythe, D. E. (1993a). Engineering knowledge: The construction of knowledge in artificial intelligence. Social Studies of Science, 23(3), pp. 445-477. https://doi.org/10.1177/0306312793023003002

Forsythe, D. E. (1993b). The construction of work in artificial intelligence. Science, Technology, & Human Values, 18(4), pp. 460-479. https://doi.org/10.1177/016224399301800404

Franklin, S., & Roberts, C. (2006). Born and made: An ethnography of preimplantation genetic diagnosis. Princeton University Press.

Gasparotto, M. (2016). Digital colonization and virtual indigeneity: Indigenous knowledge and algorithm bias. https://doi.org/doi:10.7282/T3XG9TFG

Gupta, A., & Ferguson, J. (Eds.). (1997). Culture, power, place: Explorations in critical anthropology. duke University press.

Hacking, I. (1999). The social construction of what? Harvard university press.

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. Feminist Studies, 14(3), pp. 575-599. https://www.jstor.org/stable/3178066

Hayden, C. (2007). Taking as giving: Bioscience, exchange, and the politics of benefit-sharing. Social Studies of Science, 37(5), pp. 729-758. https://doi.org/10.1177/0306312707078012

Helmreich, S. (2001). After culture: reflections on the apparition of anthropology in artificial life, a science of simulation. Cultural Anthropology, 16(4), pp. 612-627. https://www.jstor.org/stable/656650

Hobart, M. (2000). After culture: Anthropology as radical metaphysical critique. Duta Wacana University Press.

Hoeppe, G. (2015). Representing Representation. Science, Technology, & Human Values, 40, pp. 1077-1092. https://doi.org/10.1177/0162243915594025

Indiana University. (2016). "Interact" http://www.indiana.edu/~socpsy/ACT/interact.htm.

Ingold, T. (2008). When ANT meets SPIDER: Social theory for arthropods. In Material agency (pp. 209-215). Springer.

Irani, L. (2015). Justice for 'data janitors'. Public Culture, 15. https://www.publicbooks.org/justice-for-data-janitors/

Irwin, A. (2008). STS Perspectives on Scientific Governance. In The handbook of science and technology studies, 24, pp. 583.

Jasanoff, S. (Ed.). (2004). States of knowledge: the co-production of science and the social order. Routledge.

Jasanoff, S. (2016). The ethics of invention: technology and the human future. WW Norton & Company.

Johnson, D. G., & Wetmore, J. M. (2008). STS and Ethics: Implications for Engineering Ethics. In The handbook of science and technology studies, 23, pp. 567.

Kim, G. H., Trimi, S., & Chung, J. H. (2014). Big-data applications in the government sector. Communications of the ACM, 57(3), pp. 78-85. https://doi.org/10.1145/2500873

Kay, J., Barg, M., Fekete, A., Greening, T., Hollands, O., Kingston, J. H., & Crawford, K. (2000). Problem-based learning for foundation computer science courses. Computer Science Education, 10(2), pp. 109-128. https://doi.org/10.1076/0899-3408(200008)10:2;1-C;FT109

Lambek, M. (Ed.). (2010). Ordinary ethics: Anthropology, language, and action. Fordham University Press.

Latour, B., & Woolgar, S. (1986). Laboratory life: The construction of scientific facts. Princeton University Press.

Latour, B. (1991). We have never been modern. Harvard University Press.

Latour, B. (1999). Pandora's hope: essays on the reality of science studies. Harvard University Press.

Liu, Jennifer A. (2017). Situated stem cell ethics: beyond good and bad. Regenerative Medicine, 12, pp. 587-591. https://doi.org/10.2217/rme-2017-0059

Manderson, L., Davis, M., Colwell, C., & Ahlin, T. (2015). On secrecy, disclosure, the public, and the private in anthropology: an introduction to supplement 12. Current Anthropology, 56(S12), pp. 183-190. https://doi.org/10.1086/683302

Mohamed, S., Png, M. T., & Isaac, W. (2020). Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. Philosophy & Technology, pp. 1-26. https://doi.org/10.1007/s13347-020-00405-8

Morganson, V. J., Jones, M. P., & Major, D. A. (2010). Understanding women's underrepresentation in science, technology, engineering, and mathematics: The role of social coping. The Career Development Quarterly, 59(2), pp. 169-179. https://doi.org/10.1002/j.2161-0045.2010.tb00129.x

Mosemghvdlishvili, L., & Jansz, J. (2013). Negotiability of technology and its limitations: The politics of App development. Information, Communication & Society, 16(10), pp. 1596-1618. https://doi.org/10.1080/1369118X.2012.735252

Muller, Vincent C. (2014). Risks of general artificial intelligence. Journal of Experimental & Theoretical Artificial Intelligence, 3, pp. 297-301. https://doi.org/10.1080/0952813X.2014.895110

Nardi, B. (2010). My life as a night elf priest: An anthropological account of World of Warcraft. University of Michigan Press.

Nota, C. (2015). AGI Risk and Friendly AI Policy Solutions. Retrieved from https://cpnota.github.io/nota_agi_risk.pdf

Picarra, N., Giger, J.C., Pochwatko, G., and G. Goncalves. (2016). Making sense of social robots: A structural analysis of the layperson's social representation of robots. Revue europeenne de psychologie appliquee, 1-pp. 13.

Reardon, J. (2001). The human genome diversity project: a case study in coproduction. Social Studies of Science, 31(3), pp. 357-388. https://doi.org/10.1177/030631201031003002

Richardson, K. (2015). An anthropology of robots and AI: Annihilation anxiety and machines. Routledge.

Robertson, J. (2010). Gendering humanoid robots: Robo-sexism in Japan. Body & Society, 16(2), pp. 1-36. https://doi.org/10.1177/1357034X10364767

Rogers, K. B., Schröder, T., & von Scheve, C. (2014). Dissecting the sociality of emotion: A multilevel approach. Emotion Review, 6(2), pp. 124-133. https://doi.org/10.1177/1754073913503383

Scheper-Hughes, Nancy. (1995). The Primacy of the Ethical: Propositions for a Militant Anthropology. And Responses. Current Anthropology, 36, pp. 409-440. https://doi.org/10.1086/204378

Seaver, N. (2018). What should an anthropology of algorithms do? Cultural Anthropology, 33(3), pp. 375-385. http://orcid.org/0000-0002-3913-1134

Solomon, M. (2008). STS and Social Epistemology of Science. In The handbook of science and technology studies, 10, pp. 241.

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2017). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous And Intelligent Systems, Version 2. IEEE. https://standards.ieee.org/industry-connections/ec/autonomous-systems/index.html

Torrance, S. (2013). Artificial agents and the expanding ethical circle. AI & Society, 28(4), pp. 399-414. https://doi.org/10.1007/s00146-012-0422-2

Turner, Bryan S. (2007). Culture, technologies and bodies: the technological Utopia of living forever. The Editorial Board of the Sociological Review, pp. 19-36. https://doi.org/10.1111/j.1467-954X.2007.00690.x

Vanderelst, D., & Winfield, A. (2018, December). The dark side of ethical robots. In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (pp. 317-322). https://doi.org/10.1145/3278721.3278726

Veruggio, G., & Abney, K. (2011). Roboethics: The Applied Ethics for a New Science. In Robot ethics: The ethical and social implications of robotics, 22, pp. 347.

Wallach, W., Allen, C., & Smit, I. (2008). Machine morality: bottom-up and top-down approaches for modelling human moral faculties. AI & Society, 22(4), pp. 565-582. https://doi.org/10.1007/s00146-007-0099-0

Warwick, K. (2013). Artificial intelligence: the basics. Routledge.

Winchester III, W. W. (2019). Engaging the Black Ethos: Afrofuturism as a Design Lens for Inclusive Technological Innovation. Journal of Futures Studies, 24(2), pp. 55-62. https://jfsdigital.org/articles-and-essays/vol-24-no-2-december-2019/engaging-the-black-ethos-afrofuturism-as-a-design-lens-for-inclusive-technological-innovation/

Zigon, J. (2010). Moral and ethical assemblages. Anthropological Theory, 10(1-2), pp. 3-15. https://doi.org/10.1177/1463499610370520

JOURNAL OF DIGITAL SOCIAL RESEARCH

WWW.JDSR.IO ISSN: 2003-1998

VOL. 2, NO. 3, 2020, 65–83

WHAT IS DATA AND WHAT CAN IT BE USED FOR? KEY QUESTIONS IN THE AGE OF BURGEONING DATA-ESSENTIALISM

Jakob Svensson* & Oriol Poveda Guillen **

ABSTRACT

In this article we describe the rise of a data orthodoxy that we suggest to label 'data-essentialism'. We question this data-essentialism by problematizing its premises, and unveil its ideological indebtedness to deeper (previous) currents in Western thought and history. Data-essentialism is the assumption that data is the essence of basically everything, and thus provides the ideological underpinnings for the imagination of creating an Artificial Intelligence (AI) that would transform the human race and our existence. The imagination of data as an essence is in contrast to, while often conflated with, ideas of data as traces we leave behind existing in highly connected societies. This confusion over what data is, and can be used for, underlines the importance to engage in questions of the nature of data, whether everything in the universe can be described in terms of data and the implications of subscribing to such a data-essentialist worldview. We connect data-essentialism to a revival of positivism, critique a belief in the objectivity of data and that predictions based on data correlations can be fully accurate. We end the article with a discussion of how some aspects of AI rely on data-essentialist accounts and how these have a history and roots in Modernity.

Keywords: Algorithms, Artificial Intelligence, Data, Essentialism, Modernity, Positivism, Predictions

* Malmö University, Sweden.
** Freelance.

1 INTRODUCTION

Data is on the agenda today. So-called data forms the bedrock of modern policy decisions, underlies protocols of medical health, is the basis of investment strategies, informs our knowledge of the world (see Gitelman and Jackson 2013, p. 1), influences how we see ourselves and others (according to Kennedy 2016, p. 48), acts upon us (see O'Neill 2016), and thus shapes possibilities for action (according to Bowker 2013, p. 168). Hence, today there is no doubt that we are made subjects to data (see Gitelman and Jackson 2013, p. 2), determined by our *data exhausts,* an invisible ether of ones and zeroes upon which the world increasingly depends (see Jarzombek 2016, preface p.ix). With the rise of digital tech giants such as Google and Facebook, more and more aspects of our lives are mediated by their platforms as ever-increasing amounts of information are being compiled about our consumption habits, social networks and locations. According to Jarzombek (2016, preface p. x), data becomes our new *oxygen*, or should we rather say carbon *dioxide* as a growing share of our lives are dedicated to its release, capturing and processing. As hostages of these digital tech giants, we are turned into collaborators in the creation of data surpluses (see Jarzombek 2016, p. 42). But surprisingly we seem to sympathize with our captors as we participate in these practices freely (hence the title of Jarzombek's book: *Digital Stockholm Syndrome*). Because data lays out the promise of a more convenient and efficient future in which data processing algorithms know us users (customers) better than ourselves. This is nicely illustrated in a quote by an anonymous Facebook user: 'I am never quite sure if Facebook's advertising algorithms know nothing about me or more that I can admit to myself' (in Andersson Schwarz 2018, p. 68). In other words, data measuring technologies have become ingrained in the experience of the self as also the whole *Quantified Self* community is an example of.[1]

It is therefore not surprising that debates over data– how it is produced, who owns it and has access to it, and to what uses it can be put – have become key political discussions in our time. The scandal with Cambridge Analytica, browsing millions of Facebook profiles and using their data traces without consent and used for political purposes in elections, is a case in point. Still, the amount of data currently harvested and its implications for our daily lives will be negligible in comparison to what the internet of things aims to deliver in terms of all-round connectivity and data-harvesting (see Bunz and Meikle 2018). Social media giants harvesting of enormous amounts of user data (as imperative for their business models)

[1] A community of experimenters in self-tracking technologies hoping that through smarter machines and their more intimate and persistent measuring, they will reach a higher degree of self-knowing.

has awakened fears of a dystopian world in which surveillance and control by a digital 'big brother' would offer the ultimate oppressive tool for any authoritarian regime. China is already enforcing a system of mass surveillance and control using facial recognition and big data analysis technology in their so-called *Social Credit* system which are used for among other things decisions on banking credits, insurance premiums and possibilities for travelling abroad.[2] Such oppressive uses of data have led researchers to address issues of data justice, relating data-driven forms of governance to broader social justice agendas (see Dencik, Hintz and Cable 2016; O'Neill 2016; Noble 2018).

In other words, we seem to be poised at the cusp of a data revolution, which makes reflections about data – what it is and what it can be used for – all the more important. However, the nature and materiality of data is seldom attended. In order to initiate a discussion about these questions we will describe what we have observed as a rise of a data orthodoxy that we suggest to label 'data-essentialism'. This is different from the common conception of data is as *traces* we leave behind, or *exhaust* as Jarzombek (2016) would phrase it. Data-essentialism, in contrast, is based on the assumption that data is the *essence* of basically everything. An example of such data-essentialist reasoning is when acclaimed historian Harari (2015) suggests that all organisms (including humans) *consist* of data flows.[3] This idea – that the building blocks of *both* computers and organisms are data – makes the merging of life sciences and data sciences possible, providing the ideological underpinnings for the belief that the human brain can be accurately modelled in a computer (see the *Human Brain Project* funded by the European Union, https://www.humanbrainproject.eu/)[4] This opens a possibility for creating a form of AI that in the end would make the human race as we know it come to an end (for accounts of such scenarios, see Bostrom 2014; Tegmark 2017). Barriers between animals and machines collapse and the expectation is that electronic algorithms will decipher and eventually outperform biochemical ones (as Harari 2015, p. 428 argues). Harari even seems to claim that we already have the amount of data available, and the processing power, to upgrade our old algorithmic processor (i.e. the body).[5] Homo Sapiens is on the brink of evolving into a

[2] See https://en.wikipedia.org/wiki/Social_Credit_System, accessed November 27th 2019.

[3] It is unclear in his account whether the data he argues flows through human bodies are inherent to our bodies or external (or a mix of the two).

[4] Such thinking can be found in early Cybernetics in which communication and messages are considered the backbone of *both* animals and machines (see Wiener 1948). He compared the nervous system with the computing machine of his era (see p.14).

[5] We choose the term "seems" here as Harari (2015) in other parts of his book is ambivalent about the ability of technology to eventually making Homo Sapiens obsolete (see p. 458).

new species; *Homo Deus* (as is the title of Harari's book), or should we rather say *Homo Datus*?[6]

We believe that it is important to tease out how approaches to data evolve and differ in order to have an informed discussion about data and its power and politics in contemporary connected societies. Because these two views – data as essence and data as traces – are sometimes conflated. In this article we will attempt to distinguish the two by defining data-essentialism along three tenets (or beliefs) in which it differs from perceiving of data as traces; 1) that everything in the universe can be understood as data, 2) that data provides an objective picture of humans and hence 3) also may predict the future accurately. We will also critique and problematize these premises and link data-essentialism to a revival of positivism. We will end the article with unveiling its ideological indebtedness to deeper (previous) currents in Western thought and history. Accounts of superhuman AI (see Bostrom 2014; Tegmark 2017), rests not only on the assumption that humans can be reduced to data, but also on older assumptions inherited from Modernity that humans can be reduced to their minds.[7]

Rather than a coherent movement of people, data-essentialism is a way for us to illustrate how conceptions of data differ and sometimes are conflated. This confusion of what data is in contemporary accounts, became apparent when reading Cheney-Lippold's (2017) book with the rather misleading title *We are data*. Cheney-Lippold (2017) claims that we are 'made of data' (preface p. xiii) and that we are 'filled with data' (p. 3). However, it would be wrong to label Cheney-Lippold a data-essentialist. Reading the book to the end, the main message is actually that we are *not* made of data, but rather represented, categorized and regulated by data, and that data-mining and triangulating processes are increasingly automated without our direct participation. But to be acted upon by data, algorithms and automated systems, is not the same thing as to be made of data and this we will argue has important implications on what data can be used for.

2 DATA-ESSENTIALISM AND ITS THREE TENETS

While we rather subscribe to an approach to data as traces we leave behind living in societies permeated with digital technologies, data-essentialism assumes that we are made of data/outcomes of algorithmic calculations on data-flows. One example of such data-essentialist reasoning is in Harari's

[6] or *Homo Sapiens Digital* as Prensky (2009) suggests.

[7] An argument that we no longer live in Modernity but in the Global Age can be found in Poveda and Svensson 2016.

(2015) *Homo Deus*. Here he argues that human feelings are supposed to be outcomes of calculations of data in our bodies (p. 97). Free will is just biochemical processes of calculating data in order to make decisions based on probabilities (p. 328). Another example of data-essentialist thinking is Andersson's (2008) (in)famous account of the '*end of theory'*. Academics won't need theories as we have enough data and smart enough data-calculating algorithms to find patterns and hypotheses without the guidance of human thinking. Powerful computers equipped with such algorithms will be able to mine big datasets for patterns revealing effects without experimentation (as also Prensky 2009 argues), exposing patterns and relationships we didn't even know existed (see Dyche 2012), correlations that provide a full resolution of the world (see Steadman 2013), freed from human bias and framing, transcending context and thus being inherently truthful. [8] The scientist's role shifts from being proactive (suggesting theories) to reactive with algorithms doing all of the contextual work (as Steadman 2013 forecasts). This is about collecting data first and later let the algorithms ask the questions (see Croll 2012). Such data-essentialist thinking can of course be questioned. But before this we need to better understand what data-essentialist thinking consists of.

We have identified three tenets upon which data-essentialism rests and also differs from an understanding of data as traces. The first one is the belief that everything can be accurately described in terms of data flows. One example here is Harari (2015) who argues that the wall between the organic and the inorganic has been dismantled, "turning the computer revolution from a purely mechanical affair into a biological cataclysm" (p.402). He therefore concludes that the human body is a data processing system, an algorithm (see also Wiener 1948) with everything from human imagination and feelings to free will being a product of biochemical algorithms processing data in our bodies. Neurologists have convincingly argued that the brain indeed does process information from our body which then could be behind feelings, emotions and consciousness (see for example Damasio 1999). But that such information comes in the form of data (whatever data is supposed to be in these accounts), and whether its processing is following strict steps defined in an algorithmic formula, remains questionable.

The second tenet is the imagination that it would be technically possible to extract and make calculations upon the data that our bodies is supposed to consist of. This is the belief that algorithms and automated systems may arrive at insights by correlating data being extracted from us into patterns (as Brooks 2013 seems to argue) and provide a complete and objective picture of human beings as well as a full resolution on the social

[8] See also Kitchin's (2014, p. 132) critique of such faith in data.

worlds and cultures we humans organize ourselves in (see Steadman 2013). Another example is Harari (2015) who claims that with the rise of bio-metric devices (DNA scans et cetera), Google and its competitors will become an "all-knowing medical health service" (p. 392). In effect, this means that a human being can be reduced (from her bio-chemical processes to her social behaviour) to the data extracted from her in what is supposed to be a scientific and bias-free way. Others have described this as *dataism*; a "widespread belief in the objective quantification and potential tracking of all kinds of human behaviour and sociality through online media technologies" (van Dijck 2014, p. 198). This is in turn is linked to *datafication*, the paradigm for understanding sociality and social behaviour by transforming social action into online quantified data (see Cukier and Mayer-Schoenberger 2013).

This leads us to the third and final tenet of data-essentialism: By compiling and analysing increasing amounts of data harvested from human beings, it is believed to be possible to make *fully accurate predictions* about our behaviour. That data traces we leave behind can tell us great deal, seems like an uncontroversial claim. But only if we imagine data as neutral and objectively true may it allow for fully accurate predictions. Indeed, objective quantification and tracking is only possible if data is conceived of as a *neutral essence* rather than as *contextual* and *situated traces*. Taken to its logical consequence, what this third tenet postulates, is that with enough data, predictions would no longer be a matter of probabilities but would rather evolve into error-free forecasts. One example of how such thinking can have potentially harmful consequences is so-called predictive policing. While being presented as objective and bias-free, O'Neill (2016) shows how predictive policing systems send cops back to the same poor neighbourhoods, creating a toxic feedback loop since policing one street creates new data that justifies more policing in that exactly that same street. As Siegel (2013, p. 90) claims, we do not need to care about causation, explaining *the why*, when the objective is to predict the world rather than to understand it (an argument Pearl and Mackenzie 2018 refutes in their *Book of Why*).

There is no doubt that data and its processing by algorithms have wide ranging implications in terms of how we are represented, controlled and disciplined today (as O'Neill 2016, Cheney-Lippold 2017 and Noble 2018 have shown). Life in connected societies indeed increasingly takes place in and through an algorithmic media landscape processing data (as Bucher 2018 argues). We are datafied, including our friendships and relationships (see Kennedy 2016, p.10). But this is not the same as consisting of data, or that data is a neutral essence. But this begs the question of what data really is, which leads us to the next section.

3 WHAT IS DATA?

Kitchin (2014, p.18) claims that data is getting an ontological status in technology, sociology as well as biology. At the same time, he complains that little attention has been paid to data's ontological framing and the meaning of data itself (p. 25). We agree that data is often treated as ontological, but that questions about its nature and materiality remain unanswered. Cheney-Lippold (2017) does not define data despite the title of his book. Harari (2015) uses the terms *data* and *information* interchangeably without defining neither of them. This lack of definition is arguably behind confusions of what data really is, and for some researcher to treat data as a neutral essence, devoid of cultural bias.

The treatment of data as an objective and neutral reflection of reality resonates in the etymological meaning of the word as something that is given (from *datum* and the Latin verb *dare* i.e. to give, see Rosenberg 2013, p. 18). In 17th century philosophy, data equalled facts and principles that were "by agreement beyond argument" (Rosenberg 2013, p. 20). Here data is supposed to be the starting point of what we know and cannot be deconstructed. This etymological meaning is probably behind conceptions of data as an essence, "transparent, autonomous, objective and neutral" (Gitelman and Jackson 2013, pp. 2-3). However, data is not given, most often it is *captured*, extracted through observations and computation.[9] But even though the meaning of data has shifted from the rhetorical (what is beyond argument), to the observable (what can be extracted, see Rosenberg 2013, p. 36), its connotation to the objective and factual seems to have persevered. Data is supposed to have no inherent meaning (as Kitchin 2014, p. 17 argues), and therefore it has been very useful as a concept (according to Rosenberg 2013, p. 37).

There are critics of perceiving of data as objective. Data do not just exist; it has to be generated. This is nicely illustrated in Ribes and Jackson's (2013) study of the largely invisible infrastructures of data, how scientists and technicians worked hard to make data the same, comparable over a long period of time in a setting in which context and conditions were constantly changing. Etymologically it would make more sense to talk about data as *information*. Galloway (2011, p. 87) distinguishes *what is given* from *information*, meaning the act of *being formed* or *put into a form*. Hence, the data that is most often referred to today, is *computable* data, data that is made 'algorithm-ready' (Bucher 2018, p. 5), 'scrubbed' (Gitelman and Jackson 2013, p. 7) and 'cleaned' (Kennedy 2016, p.108) for computer algorithms to use in their calculations. And computer-readied data is not

[9] Which leads Kitchin 2014, page 2, to suggest we should rather talk about *capta* rather than *data* (from Latin *capere* i.e. to capture).

formless. It is captured when being measured/collected, a capture which shapes the data (see Ribes and Jackson 2013) and put into a quantified form of ones and zeroes in order for computers to process it (see Kennedy 2016, p. 10). Data is thus always dependent on developments around its capturing and scrubbing (Pink et al. 2018).

This suggests that data is deeply cultural and infused with societal norms and values. Data does not naturally appear as it is collected and manipulated by people, shaped by human decisions, interpretations and filters (see Kennedy 2016, p. 110; Cheney-Lippold 2017, p. preface xiii). Behind data production there are assemblages of people, places, documents, practices and technologies, making data a product of complex processes in order to be useful for the contexts in which it appears (as Ribes and Jackson 2013 show). Krippendorf (2016) therefore defines data as a *human artifact*. Indeed, data is both social (situated in a context), material in that it has a form. In terms of computer data this would be in the form of bits stored on a hard drive, and depending on infrastructures (such as data centres and cables, see Holt and Vonderau 2015). Raw data is thus 'an oxymoron' (as Gitelman and Jackson 2013 argue) and should be 'cooked with care' (Bowker 2005, p. 184), otherwise it might 'rot' (Boellstorff 2013) and thus be in need of 'repair' (Pink et al. 2018).

This reasoning above is surprisingly uncontroversial. There is even a field called *critical data studies* (see Illiadis and Russo 2016). In tech literature such as *Algorithms for Dummies* (Mueller and Massaron 2017) it is clearly stated that data is not raw, it is managed and that programmers and algorithms are so-called 'data managers' (see p. 68). Once we take away the neutrality and objectivity of data, admit that it is socio-cultural, the question is if the premises upon which ideas of superhumans and AI rest also would start to unravel? Because, if we agree that data is a human construct, how could everything in the universe be described in terms of data? Or is data-essentialism an extreme form of social constructionism?

Here it seems that data-essentialism is connected to the hype around so-called *big data* (see Cukier and Mayer-Schoenberger 2013). This is the imagination that large datasets open the possibility for a higher form of intelligence and knowledge and thus may generate insights that were previously unavailable through hidden patterns and correlations in data points (i.e. data-essentialism's 2nd tenet). Andrejevic (2020, p. 35) talks about this as a *fantasy of framelessness*. Automated collection and processing of data is thought of as final and ultimate, as it nurtures *a fantasy of total information collection* (Andrejevic 2020, p. 35) out of which decision untouched by human prejudices can be made.

Big data is the outcome of an increasing ease and thus intensification of data collection and storage coupled with computers with increased processing power. Digital storage solutions have reduced the cost and space of retaining data, and the networking of computers has facilitated the transfer and sharing of data (see Kitchin 2014, pp. 31, 82). However, big data is a relative term. Big data is only big in relation to previous amount of data collection and processing. It is indeed big compared to what human beings alone can process, but it is small compared to the amount of data potentially available (see Poveda and Svensson 2016). It is therefore important not to confuse *big* data with *all* data (as also Andrejevic 2020 argues). Data harvested through measurement is always a selection from the total sum of all possible data (see also Kitchin 2014, p. 3). And since so-called big data cannot capture the whole picture (it is always framed), calculations on big data sets are biased from the beginning as they constitute partial orders, localized totalities and with an ability to only gaze in some directions but not others (see Kitchin 2014, p. 133). As Cukier and Mayer-Schoenberger (2013) reminds us of, “however dazzling the power of big data appears, its seductive glimmer must never blind us to its inherent imperfections” (p. 28).

4 WHAT CAN DATA BE USED FOR?

If we agree that data is (inter)subjective, infused by socio-cultural norms and values (at least in part), we should also start to ask what it can be used for. In an interview with a software engineer he stated that “data you do not do anything with, is uninteresting”, that ”data can be bad and not useful” and that “data only treats one part of reality” (in Svensson 2020). Hence, if we know that data from the beginning is biased, that big data is far from all data, how can the predictions it makes be fully accurate and applicable? Furthermore, algorithms are trained to find correlations in data, make associations and construct patterns and out of these make predictions out of probabilities. Patterns need big numbers and thus mostly work on big data sets. It is by collecting *enough* data that not only the past and present are mapped, but also the future. And the more the coin is flipped, the more the result will converge upon the precalculated probability (see Steiner 2012). Indeed, patterns are all about prediction which is all about probabilities. Already Wiener (1947, p. 34) was occupied with the ability to predict out of information. This fascination with prediction goes all the way back to Leibniz who thought humans were programmed to behave in certain manners (according to Steiner 2013, p. 61). But correlation does not supersede causation and data does not understand causes and effects (as Pearl and Mackenzie 2018, p. 21 argue). As Cukier and Mayer-Schoenberger (2013) states, the use of big data might imply we will need to give up our quest to discover the cause of things. Looking for patterns might help predict the future, answer to what probably (but not certainly) will happen, but not why this will happen. Hence, predictions are only probabilities and are not always correct. And as Pearl and Mackenzie (2018, p.47) argue, causation is not reducible to probabilities. Even if predictions would be based on completely neutral and correct data, the people using these systems might not be, as the case of predictive policing has shown. It becomes dangerous if we treat predictions of probabilities as undeniable truths.

Since algorithms will not ask why they get the results they get or what the consequences of their results might be, it makes them blind to ethical issues (see Diakopolous 2016). This is about outsourcing the ordering of the world we inhabit to algorithms lacking reflexive capabilities and lacking agency to handle the messiness of the present (see Klinger and Svensson 2018). Hence, there are numerous examples of when algorithms fail, such as Amazon being accused of homophobia (see Striphas 2015), Google of racism (see Noble 2018), gender biases of image-search algorithms (see Kay, Matuszek and Munson 2015) and cases where black people are not recognized as humans in face-recognition algorithms (see Sandvig et al. 2016).

It is only if we believe in the objectivity of data, imagine that it would be technically possible to extract and make calculations upon the data in our bodies, that patterns found in big datasets could be used for fully accurate predictions. But if we believe that data are traces that we leave behind in a digital existence, such predictions would always be based in the past. This contemporary craving for patterns may have dire consequences when making judgements about people's ability to change destructive patterns of the past (see O'Neill 2016). The past is not necessarily determinative of the future, people can change. If we instead approach data as traces from past behaviour online, algorithmically calculated patterns, these cannot be believed to predict the future with 100 percent of accuracy. If there is something we have learned in the history of humankind, is that it has taken many unexpected turns.

To be human is to be random, unfinished, imperfect and disorderly, to be a constant "beta version" (as Cheney-Lippold 2017, p. 90, eloquently puts it). At the same time, most of data analytics and processing are about orderliness, calculations and finding patterns which are supposed to predict future behaviour. A software engineer interviewed actually described code as a grammar with no exceptions (see Svensson 2020). This was the reason why he loved coding, comparing this to struggles with German grammar at high school. But as humans we have plenty of exceptions and at times we act randomly and in a surprising manner. As Morozov (2013, introduction p. xiii) argues, sometimes imperfect is good enough and even much better than perfect. It seems that the orderliness of programming and code languages, are at odds with human imperfectness and randomness. Maybe some things are just un-representable by computer-readied data in the form of ones and zeros (as also Galloway 2011 argues). Maybe this is why we sometimes feel creeped out by our datafied selves (see Cheney-Lippold 2017, p. 193). We are recognizable but in an odd way. It becomes uncanny in the same way that robots can be creepily similar, but not quite like the real thing (the so-called uncanny valley).[10] Behind the perfect surface, there is just mechanical impulses. Digital computers can mimic the actions of human behaviour as already Turing (1950, p. 437) forecasted. But is an imitation the real thing? Arguably what is missing in our datafied replicas/upgrades is irrationality and randomness, patterns and also correlations, but with plenty of exceptions.

5 A REVIVAL OF POSITIVISM?

The belief that data can capture everything with full resolution, freed from human bias, framing and context does ring a bell. The bringing of the unruly social world into the formal study of the natural sciences, rendering culture and society computable is surrounded by a discourse of positivistic measurement. It thus seems data-essentialism is accompanied with a revival of positivism within the Social Sciences. Indeed, as Kitchin (2014, pp. 139-140) argues, data-driven sciences favour transforming research about humans and their societies to something resembling natural and engineering sciences, offering opportunities for a 'truthful' study of human life. Törnberg (2019) labels the use of API-based technologies to inductively seek patterns as *predicative positivism.* Indeed, datafication implies transforming sociality, behaviour and culture into *quantified data* to be used for real-time tracking and predictive analysis.

Following this discourse of positivistic measurement, Anderson (2008) has (in)famously argued that theory has come to an end, and that we now have enough data and fast enough computers to actually study *the physics of culture*. He thus seems to suggest that data will be able to speak for itself. This can of course be questioned (see also Törnberg 2019). Bucher (2018, p. 24) for example claims that without algorithms, data would just flow without any particular direction. Algorithms are actually an outcome of media logics rather than a replacement of them (see Klinger and Svensson 2018 for an outline of this argument). Algorithms are based on hypotheses from the beginning (see Bucher 2018, p. 25). And even if data harvested from social media platforms is supposed to reflect human behaviour, the algorithms employed (by Google, Facebook and others) are intrinsically selective and manipulative to suit the interests of these companies (see van Dijck 2014, p. 200). Hence, it is easy to dismiss statements such as those claiming that data speaking for itself. But it is

[10] See https://en.wikipedia.org/wiki/Uncanny_valley, accessed August 21th 2020.

important to understand that one of the strongest epistemic conditions shaping data imaginaries today, is the self-evidence of numbers. Data's connection to numbers and mathematical functions, gives it an allure of neutrality and objectivity, which in turn makes humans look particularly subjective and biased in comparison (see Bucher 2018, p. 56).

Kennedy (2016, p. 150) talks about a 'pervasive desire for numbers' as an emerging rationality today. She shows in her studies of public sector organizations that mere numbers are met with enthusiasm (even though it was not always clear what they stood for). Kennedy connects this desire to earlier studies about trust in numbers that seem to support the prestige and power of quantitative methods. Numbers can be understood from far away and are universal as they can be shared across cultures (see Kennedy 2016, p. 81). They are impersonal, therefore also appear to be objective and thus credible. This to the point that even friendships and sociality are quantified in a positivistic manner of objective measurement (see Bucher 2018, p. 9). What was once qualitative has been turned into numbers. According to Kennedy (2016, pp. 100-101), this limits the possibility to discuss the ways in which data is made and shaped.

This desire for numbers, with its allure of objectivity and neutrality, is accompanied with a belief of unbiased calculation, the translation of everything into mathematical symbolic language following mathematical laws. Algorithms introduce and privilege quantification and automation, the ordering of various types, statistical reasoning and large numbers (see Bucher 2018, pp. 31-32). And if we are made up of data and our bodies are just bio-chemical algorithms processing this data, this also means that we humans could be fully predicted in mathematical formulas, that the entirety of our everyday life practices and ourselves are subject to – and constituted by – perpetual calculation (as Raley 2013, p. 126 argues). Harari (2015, p. 99) gives the example of a baboon spotting some ripe bananas in-between him and a lion. His body will calculate how hungry he is together with probability of success, which will then result in a feeling of bravery or caution. In other words, sensations, emotions and actions are a result of mathematical calculations on the data inside of us according to Harari (2015, p. 124). Harari (2015, p. 101) even argues that attraction and beauty are results of years of calculating data about reproduction with successful offspring. But is it really possible to reduce subjective and intersubjective experiences such as beauty to mathematical calculations on data? If it is one thing we know about beauty, it is that it is culture specific, whereas today's Western beauty ideal of female skinniness is not related to being successful at birth-giving (arguably it is the other way around). Indeed, as Bucher (2018, p. 11) puts it, by reducing human connections to algorithmic calculations, we risk dehumanizing sociality. People are not a math problem, and people are more complicated than an equation, more complex and unpredictable than what can be broken down into a few steps of instructions in a computer algorithm (as Bucher 2018, pp. 104-105, argues).

6 AI AND ITS ROOTS IN MODERNITY

By reducing us humans, our connections and our behaviour to data being algorithmically processed, calculated by our bodies or/and computers, data-essentialism provides the ideological underpinnings for the belief that humans can be replaced by AI with far greater capabilities (see for example Bostrom 2014; Harari 2015; Tegmark 2017). According to this line of reasoning, it would be technically possible to create machines that are better and more efficient at processing our data. This claim is currently challenged by science's poor understanding of how human consciousness works (see Damasio 1999). But this might be a temporary obstacle that new research perhaps could contribute to overcome.

A more problematic objection can be found in AI's understanding of the human. It is worthwhile to interrogate in which ways the reduction of being human to data is indebted to older forms of reductionism. In religious thought, the search for a human essence detached from the physical body led to the notion of an immortal soul. In modern times, Descartes (2017) gave scientific sanction to the body/soul dualism previously upheld by theologists by reframing it as the body/mind split. Descartes, too, conceived of bodies as machines. Data-essentialism reproduces in a magnified fashion the soul/body controversy in Christianity. The project to de-incarnate the human and retrieve her essence, has ancient roots but current discussions around AI seem not to account for this ideological lineage and presents it as novel, what is in fact a cultural bias with a long history in Western thinking.

It is worthwhile to look pass the hype that surrounds AI and to question its claim for novelty. As a matter of fact, data-essentialisms' first and second tenets were already expressed by Weber (2008) in his famous lecture series when he described *disenchantment* as "the knowledge or belief that if we only wanted to, we could learn at any time that there are, in principle, no mysterious unpredictable forces in play, but that *all things— in principle— can be controlled through calculation*" (p. 35, emphasis in the original). The third tenet of data-essentialism, the belief that with enough data it would be possible to make fully accurate predictions, seems also to be behind Weber's reference to absolute control. As much as data-essentialism toys with the idea of rendering humans obsolete, it is important to underline that, historically speaking, the modern project of human mastery lies at its core. The modern belief in endless progress lurks behind ideas upgrading humans with computer technology. Indeed, As Morozov (2013, introduction p. ix) argues, to question Silicon Valley's quest to solve any kind of problems with tech, has become equivalent of questioning Enlightenment itself. Also, Rosenberg (2013, p. 15) associates the rise of the concept of data to Modernity and Jarzombek (2016, p. 39) argues that data processing is about making the Self and Others predictable, identifiable and exploitable. To participate in the project of Modernity has always meant that one becomes "a calculable subject" (according to Raley 2013, p.126). And what is the meaning of AI apart from *progress* and a trust in an upgraded future? However, even if we would consist of data flows, it would still be uncertain that we would process this data in a *rational* manner. The modern belief in rationality, that human beings act (at least in the aggregate) as rational beings and in their self-interest is part of AI. However, global warming clearly shows otherwise. For the sake of ourselves and our survival, the most rational thing to do would be to reduce our carbon footprints (while on the contrary, it seems to be increasing). Indeed, the façade that attempts to present AI as a dispassionate reckoning with the objective realities of today, our data and algorithm saturated world belies a much more complicated and problematic genealogy of its foundational principles and ideas.

Finally, it is relevant to point out that our critique of data-essentialism is not predicated upon any form of human exceptionalism. Intelligence and a rich emotional life are not an exclusive prerogative of human animals. Our critique aims rather at problematizing the premises of data-essentialism and to unveil its ideological indebtedness to deeper currents in Western thought and history that have little to do with claims of objectivity and neutrality.

7 CONCLUDING REMARKS

Many people today believe in data as we ask Google and Facebook for advices on a range of different matters. Contemporary life is indeed characterized by data collection and processing. As we are *thrown* into a digital existence (see Lagerkvist 2017), digital tech giants and data scientists are increasingly powerful centres around which our existence gravitate. But acknowledging the importance of data, conceiving of data as contextual and situated traces we leave behind in an increasingly computer saturated world is substantially different from reducing our existence and bodies to data. As we have discussed in this article, such data-essentialism is indebted to modernist thinking about progress, calculation and rationality.

Harari (2015, p. 207) does emphasize the role of fiction for societies to function. The importance of SciFi (Science Fiction) in tech in general and AI in particular cannot be understated. SciFi aesthetics, with its connection to futurism, are all over tech culture (see Svensson 2020). The modern imagination of a disembodied future also resonates in SciFi classics such as

Gibson's (1984) Neuromancer. At the 2019 South by Southwest festival Cassie Kozyrkov, chief data scientist at Google, argued that the only reason AI got funding in its early days was because of its appeal to SciFi. Similarly, data-essentialism seems to be based on a powerful modern fiction of humans as rational, predictable and therefore also controllable.

Bowker (2013 p. 171) writes that computers may have data, but that not everything in the world is given. Indeed, it makes more sense to understand data as partial translations (as every translation is partial, imperfect) of perceived reality in mathematical language. As such, data-essentialism seem to suffer from a poor understanding of semiotics as they mix up the sign with the thing itself. In this sense, data-essentialism is social constructionism trapped inside a cage of mathematical language, which, by virtue of being more abstract than regular human language, appears to be purer or even divinely inspired (as in Harari' 2015 account of *Homo Deus*). Data is not only a representation; it is also always a sample. Even big data is only a representation, not a totality, stand-ins for phenomena of theoretical or practical importance (see Krippendorf 2016). And to base our whole being, existence and future on partial data-traces we leave behind in mathematical language, on the "residues of human existence in a digital world" (Cheney-Lippold 2017, p. 89), would be akin to a *synecdoche*, to take a small piece and make it a representative sign of a totality.

As AI is developing now, there is no reason to believe it can fully replicate humans any time soon. Today AI is only executive while humans also think creatively and have a reflective character (see Hindi 2017). Data processing machines can show emotions but not feel them and this is different. Even a data-enthusiasts such as Domingos (2015) state that only because computers can learn "they will not magically acquire a will of their own" (p. 45). Case (2018) therefore argues that humans *together with* AI (something Case labels as *centaurs*) seem to be a winning team (even against teams of computers only). So, it seems that intelligence is not a single dimension, and that human intelligence includes random, creative, unruly and scattered elements that are hard to capture in algorithms processing readied/cleaned data.

Turing (1950, p. 440) with his focus on imitation and mimic, suggested a clear hierarchy from the human to the machine. As a gay man in the UK during the World War 2, he knew what it was like having to pass as a straight man. Today transgender activist Vanessa López raises questions about what it takes to pass as a woman in a Western society (see her book from 2014 about her regretting her gender reassignment surgery). In a similar manner we could ask whether we cannot let artificial intelligence be artificial intelligence? Does it have to *pass* as human? Why this pre-occupation with *passing* within AI? We should instead focus on what machines and AI are good at and what humans are good at, and how we together can be at the service in relation to the big problems we as humans and our planet are facing, such as xenophobia, polarization, intolerance and climate change.

FUNDING STATEMENT (AND ACKNOWLEDGMENTS)

We want to recognize the generous funding from the Swedish Research Council as well as the fellowship program at Weizenbaum Institute for the Networked Society, making this research possible. Furthermore, we would like to acknowledge colleagues at Malmö University providing important feedback to our ideas at an early stage, as well as this Journal's excellent reviewers.

REFERENCES

Anderson, C. (2008). End of Theory. The Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine*, 23 June 23. https://www.wired.com/2008/06/pb-theory/, accessed 16 July 2016

Andersson Schwarz, J. (2018). Umwelt and Individuation: Digital Signals and Technical Being. In Lagerkvist, A. (ed), *Digital Existence. Onthology, Ethics and Transcendence in Digital Culture*. New York: Routledge, pp. 61-80

Andrejevic, M. (2020). *Automated Media*. London Routledge

Bowker, G. C. (2005). *Memory Practices in the Sciences*. Cambridge: MIT Press

Bowker, G. C. (2013). Data Flakes: An afterword to "Raw Data" Is and Oxymoron. In Gitelman L. (ed.), *Raw Data is an Oxymoron.* Cambridge: MIT Press, pp. 167-171

Boellstorff, T. (2013). Making big data, in theory, *First Monday*, 18(10). https://firstmonday.org/article/view/4869/3750, accessed 8 December 2017

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press

Brooks, D. (2013). The Philosophy of Data, *New York Times*. 8 February 18. https://www.nytimes.com/2013/02/05/opinion/brooks-the-philosophy-of-data.html, accessed 20 August 2017.

Bucher, T. (2018). *If … Then. Algorithmic Power and Politics*. Oxford: Oxford University Press.

Bunz, M., and Meikle, G. (2018). *The Internet of Things*. Cambridge: Polity Press.

Case, N. (2018). How to become a centaur. *Journal of Design & Science,* https://jods.mitpress.mit.edu/pub/issue3-case, accessed 20 August 2017

Cheney-Lippold, J. (2017). *We are Data. Algorithms and the Making of our Digital Selves*. New York: New York University Press

Croll, A. (2012). Big data is our generation's civil rights issue and we don't know it, *O'Reilly Radar*, 2 August 2012. https://www.cc.gatech.edu/~beki/cs4001/big-data.pdf, accessed 20 August 2017

Cukier, K., and Mayer-Schoenberger, V. (2013). The Rise of Big Data. How It's Changing the Way We Think about the world, *Foreign Affairs,* 92 (3), pp. 28-40

Damasio, A. (1999). *The Feeling of What Happens. Body and Emotion in the making of consciousness*. New York: Houghton Mifflin Harcourt Publishing Company

Dencik, L., Hintz, A., and Cable, J. (2016). Towards data justice? The ambiguity of anti-surveillance resistance in political activism, *Big Data & Society*, July -December, pp. 1-12. https://journals.sagepub.com/doi/10.1177/2053951716679678, accessed 20 August 2017

Diakopoulos, N. (2016). Accountability in Algorithmic Decision Making. *Communications of the ACM*, 59(2), pp. 56-62

Descartes, R. (2017). *Meditations on First Philosophy, with Selections from the Objections and Replies*. Translated and edited by John Cottingham. 2nd edition. Cambridge: Cambridge University Press. Original work published 1647

Domingos, P. (2015). *The Master Algorithm. How the Quest for the Ultimate Learning Machine will Remake our World*. London: Penguin

Dyche, J. (2012). Big Data Eurekas do not just happen. *Harvard Business review blog*, 20 November 2012. https://hbr.org/2012/11/eureka-doesnt-just-happen, accessed 20 August 2017

Galloway, A. (2011). Are Some Things Unrepresentable? *Theory, Culture & Society*, 28(7-8), pp. 85-102

Gibson, W. (1984). *Neuromancer.* London: Gollancz

Gitelman, L., and Jackson, V. (2013). Introduction, In Gitelman L. (ed.), Raw Data is an Oxymoron. Cambridge: MIT Press, pp. 1-14

Harari, Y.N. (2015). *Homo Deus. A Brief History of Tomorrow*. London: Vintage Books

Hindi, R. (2017). Will Robots take over? *the Conference*, Malmö august 2017. https://urplay.se/program/202921-ur-samtiden-the-conference-2017-kommer-robotarna-ta-over, accessed 15 December 2017

Holt, J., and Vonderau, P. (2015). "Where the Internet Lives": Data Centers as Cloud Infrastructures. In Parks, L., and Starosielski, N. (eds), *Signal Traffic. Critical Studies of Media Infrastructures*. Chicago: University of Illinois Press, pp. 71-93

Illiadis, A., and Russo, F. (2016). Critical Data Studies. An Introduction, *Big Data & Society*, July-December, pp. 1-7. https://journals.sagepub.com/doi/abs/10.1177/2053951716674238, accessed 20 August 2017

Jarzombek, M. (2016). *Digital Stockholm Syndrome in the Post-Ontologicsl Age.* Minneapolis: University of Minnesota Press

Kay, M., Matuszek, C., and Munson, S.A. (2015). Unequal representation and gender stereotypes in image search results for occupations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI 15). New York: ACM, pp. 3819-3828

Kennedy, H. (2016). *Post, Mine, Repeat. Social media data mining becomes ordinary*. London: Palgrave Macmillan

Kitchin, R. (2014). *The Data Revolution. Big Data, Open Data, Data Infrastructures & Their Consequences*. London: Sage

Klinger, U., and Svensson, J. (2018). The End of Media Logics? On Algorithms and Agency. *New Media & Society,* 20(12), pp. 4653–4670

Krippendorf, K. (2016). Data. In *International Encyclopedia of Communication Theory and Philosophy*. New York: Wiley. https://onlinelibrary.wiley.com/doi/10.1002/9781118766804.wbiect104 , accessed 15 December 2018

Lagerkvist, A. (2017). Existential media: Toward a theorization of digital thrownness. *New Media & Society*, 19(1), pp. 96–110

López, V. (2014). *Jag har ångrat mig* (I have changed my mind) Stockholm: Two-Spirit Publishers

Morozov, E. (2013). To Save everything, Click here. The folly of Technological Solutionism. New York: Public affair

Mueller, J.P., and Massaron L. (2017). *Algorithms for Dummies.* Hoboken: John Wiley and Sons

Noble, S.U. (2018). *Algorithms of Oppression. How Search Engines Reinforce Racism.* New York: New York University Press

O'Neill, C. (2016). *Weapons of Math Destruction.* New York: Crown Publishing

Pearl, J., and Mackenzie D. (2018). *The Book of Why. The New Science of Cause and Effect.* London: Pengiun Books

Pink, S., Ruckenstein, M., William R., and Duque, M. (2018). Broken data. Conceptualising data in an emerging world, *Big Data & Society*, January-June, pp. 1-13. https://journals.sagepub.com/doi/full/10.1177/2053951717753228, accessed 8 December 2018

Poveda, O., and Svensson, J. (2016). Re-thinking the Global Age as Interdependence, Opacity and Inertia. *Triple C,* 14(2), pp. 475-495.

Prensky, M. (2009). H. Sapiens Digital: From Digital Immigrants and Digital Natives to Digital Wisdom. *Innovate: Journal of Online*

Education, 5(3). https://nsuworks.nova.edu/cgi/viewcontent.cgi?article=1020&context=innovate, accessed 8 December 2018

Raley, R. (2013). Dataveillance and Countervailance. In Gitelman L. (ed.), *Raw Data is an Oxymoron*. Cambridge: MIT Press, pp. 121-145.

Ribes, D., and Jackson, S.J. (2013). Data Bite Mam: The Work of Sustaining a Long-Term Study. In Gitelman L. (ed.), *Raw Data is an Oxymoron.* Cambridge: MIT Press, pp. 147-166.

Rosenberg, D. (2013). Data before the Fact. In Gitelman L. (ed.), *Raw Data is an Oxymoron*. Cambridge: MIT Press, pp. 15-40

Sandvig, C., Hamilton, K., Karahalios, K., and Langbort, C. (2016). When the Algorithm itself Is a Racist. Diagnosing Ethical Harm in the Basic Components of Software, *International Journal of Communication*, 10, pp. 4972-4990

Siegel, E. (2013). *Predictive Analytics.* Hoboken: Wiley

Steadman, I. (2013). Big data and the death of the theorist. *Wired Magazine*, 25 January, https://www.wired.co.uk/article/big-data-end-of-theory, accessed 8 December 2018.

Steiner, C. (2012). *Automate this. How algorithms came to rule our world.* New York: Penguin Books

Striphas, T. (2015). Algorithmic Culture. *European Journal of Cultural Studies*, 18(4-5), pp. 395-412

Svensson, J (2020). *Wizards of the Web. A journey into tech culture, mathemagics and the logics of programming*. Göteborg: Nordicom (forthcoming)

Tegmark, M. (2017). *Life 3.0. Being Human in the Age of Artifical Intelligence.* New York: Vintage Books.

Turing, A.M (1950). Computing Machinery and Intelligence. *Mind*, 49, pp. 433-460

Törnberg, A. (2019). Teorins död? Om framväxten av en digital empirism. *Fronesis,* 64-65, pp. 132-146.

Van Dijck, J. (2014). Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology. *Surveillance and Society,* 12(2), pp. 197-208

Weber, M. (2008). Science and Vocation. In Dreijmanis, J. (ed), *Weber's complete writings on academic and political vocations*. New York, NY: Algora Publishing, pp. 1917-1919

Wiener, N. (1948). *Cybernetics, or Control and Communication in the Animal and the Machine*. Cambridge: MIT Press

JOURNAL OF DIGITAL SOCIAL RESEARCH

WWW.JDSR.IO ISSN: 2003-1998

VOL. 2, NO. 3, 2020, 84–125

PRACTICAL AI TRANSPARENCY: REVEALING DATAFICATION AND ALGORITHMIC IDENTITIES

Ana Pop Stefanija & Jo Pierson*

ABSTRACT

How does one do research on algorithms and their outputs when confronted with the inherent algorithmic opacity and black box-ness as well as with the limitations of API-based research and the data access gaps imposed by platforms' gate-keeping practices? This article outlines the methodological steps we undertook to manoeuvre around the above-mentioned obstacles. It is a "byproduct" of our investigation into datafication and the way how algorithmic identities are being produced for personalisation, ad delivery and recommendation. Following Paßmann and Boersma's (2017) suggestion for pursuing "practical transparency" and focusing on particular actors, we experiment with different avenues of research. We develop and employ an approach of *letting the platforms speak* and *making the platforms speak*. In doing so, we also use non-traditional research tools, such as transparency and regulatory tools, and repurpose them as *objects of/for study*. Empirically testing the applicability of this integrated approach, we elaborate on the possibilities it offers for the study of algorithmic systems, while being aware and cognizant of its limitations and shortcomings.

Keywords: datafication; algorithmic identity; practical transparency; methodology; digital methods; subject access request.

* imec-SMIT, Vrije Universiteit Brussel, Belgium.

1 INTRODUCTION

Today, there is almost no area in everyday life that has not been mediated or impacted by Artificial Intelligence (AI). From recommender systems for news, apps, routes, products, to job applications, financial services, health care, education, criminal justice, etc., individuals have been increasingly, to lesser or greater degree, subjected to the automated decision making (ADM) by some kind of algorithmic and AI systems. More and more decisions impacting individuals are based on what we can call 'algorithmic identity' (Cheney-Lippold, 2011; but also Jarrett, 2014; Reigeluth, 2014) — guided by extensive profiles about people, uncovering their affinities and interests and predicting their behaviour. With the ubiquity of these ADM and AI systems, it becomes an issue of urgency to be able to investigate them, reveal their workings, and explain their outputs and impact.

We could say that this article is a "byproduct" of our attempt to investigate how algorithmic identities are being produced by few sampled actors (*Facebook, Google, Quantcast* and *Oracle)* participating in the process of algorithmic identity building for personalisation, ad delivery and recommendation. For us that meant a critical investigation of the inner workings of algorithmic systems, of the datafication practices that enable the algorithmic identity creation, in particular the actors participating in these processes, the types of data that are used, the sources of data and — importantly — their relation to the inference-making processes that are building blocks for the algorithmic identities. But an analytical inquiry like this confronted us with the question of how to investigate these issues? How does one do research on algorithms and their outputs when confronted with the inherent algorithmic opacity and black box-ness as well as with the limitations of API-based research and the data access gaps imposed by platforms' gate-keeping practices? How does one overcome and manoeuvre the limitations when dealing with data provided or extracted from these platforms while being aware of and critical of the 'methodological bias' (Marres and Gerlitz, 2015), the un-rawness of data (Gitelman, 2013) and the level of mediation (van Es et al., 2013)? This had led to our research focus of investigating the process of how an individual algorithmic is being created by few different platforms. To achieve this, we experiment with novel avenues for methodologically investigating this process and the underlying datafication processes and practices. In doing so, we also contribute towards answering the question of how to uncover and explain datafication and algorithmic identity.

This paper documents and discusses our attempts and experimentations in studying and researching algorithms while doing digital social research (Lindgren, 2019). We rely on data collection methods that manoeuvre around the API restrictions and make use of non-traditional data sources, like transparency and regulatory tools. We do this by using a mixed method design for which we developed and adopted two approaches: *letting the platforms speak* and *making the platforms speak*. This led to an investigation on two levels, interface and software, while employing two corresponding overarching methods, technography and digital methods (Rogers, 2017). Experimenting with this methodological setup, our experience and results show that, while there are limitations, this approach enables an in-depth critical inquiry and generates valuable new insights into the processes of algorithmic construction of identity, data extraction and inferential analytics.

What follows is an outline of the techniques we employed around the limitations we encountered when dealing with platforms' algorithmic systems for the purpose of research. We see this approach as just one possible path for doing research *on* algorithms, AI and platforms. As such, it does not aim to be taken as a generalizable, "apply-to-all" approach, but it aims foremost to inspire, test and experiment, and explore the possibilities and limits of different approaches and tools. We will elaborate on the rationale behind the approach, the methods chosen, the particular tools and research protocols applied, as well as the specific steps taken. First, we discuss some recent developments in digital social research, the restrictions imposed by platforms and the very nature of algorithms, and elaborate how that impacts the ability to do digital social research. This is followed by outlining our methodological design and rationale and detailing the particular steps and tools we used. A substantial part is dedicated to the elaboration of our results. We conclude with a discussion on the advantages and limitations of the particular methodological choices and tools.

2 PRACTICAL AI TRANSPARENCY

The networked infrastructure of the internet, with its technological capacity to track user movements across different web sites, apps and servers, has given rise to an industry of web analytics firms that are actively amassing information on individuals and fine-tuning computer algorithms to make sense of that data. Via the process of *datafication* – 'the transformation of the social actions of their users to quantified data' (Mayer-Schönberger and Cukier, 2013, p. 78) – and the collection of data via tracking technologies, combined with the analytics capabilities of algorithms and companies, the aim of many of these companies it to create what Cheney-Lippold (2011) calls 'algorithmic identity' – "an identity formation that works through mathematical algorithms to infer categories of identity on otherwise anonymous beings" (p.165). Datafication can be understood both as *action* and as *aim*. As an action, it means "transformation of social action into online quantified data, thus allowing for real-time tracking and predictive analysis" (van Dijck, 2014, p.198). As an aim, it relates to the pursuit to collect, monitor, analyse, understand and use people's behaviour for behaviour prediction, affinity profiling, but also for 'unstated preset purposes' (van Dijck, 2014, p.205). Raley (2013) calls the latter 'data speculation', i.e. a value yet to be added to the data and 'informational patterns still to come' (p. 123). This is closely tied, first, with the belief "in the objective quantification and potential tracking of all kinds of human behaviour and sociality through online media technologies" (ibid., p.198) to which van Dijck refers to as 'dataism', and second, with the 'collect everything' approach (van Dijck, 2014; Sadowski, 2019; Andrejevic & Gates, 2014). The creation of an algorithmic identity is possible because of the process of datafication, and datafication and dataism are the building blocks for behaviour prediction and affinity profiling, among other things, for targeted advertising and personalisation. However, this algorithmic identity is a construct, "it is not the personal identity of the embodied individual but rather the actuarial or categorical profile of the collective which is of foremost concern' to new, unenclosed surveillance networks" (Hier, 2003: 402 in Cheney-Lippold, 2011, p. 177). So how do we investigate the process of algorithmic identity and the underlying processes of datafication?

In his recent article Axel Bruns (2019) (rightfully) states that the APIcalypse has arrived and it seriously impacts our ability as social science researchers to critically study society via the digital. This results in a restriction as regards who has access to the platforms' data via their Application Programming Interfaces (APIs), which makes access to data either impossible or possible only for the chosen few and under strict conditions. Hence the APIcalypse limits the possibilities to inspect and investigate phenomena happening "in the digital". The importance of this gatekeeping is even greater if we consider that the online is never a separate realm, as decisions made about individuals based on the digital traces they leave behind can impact their offline lives as well. Being severely restricted and limited in investigating the digital and the algorithmic, seriously impacts the ability of researchers and scientists to investigate and criticise these systems, hold to account their proprietors and remedy their outputs.

To borrow the definition by Venturini and Rogers (2019), API based research is

> an approach to computational social sciences and digital sociology based on the extraction of records from the datasets made available by online platforms through their application programming interfaces (or APIs). This type of research has allowed the collection of detailed information on large populations, thereby effectively challenging the distinction between qualitative and quantitative methods. (p. 1).

As such, this approach enabled the studying of a variety of phenomena pertaining to the interplay and mutual influence of both technology and society, and mediated numerous findings, previously not possible at such a large scale. However, this is not the only approach undertaken or proposed by researchers and scholars for studying the digital, or the best one. As Venturini (2018) notes, 'when all you have is a Twitter feed, everything looks like a hashtag' (p. 4210). This refers to the limitations imposed by the affordances of the platforms when we see them as objects of research. We use these statements as an entry point to discuss some of the approaches developed for studying the relationship between the digital, the algorithmic and the societal, their limitations and shortcomings. In the paragraphs that follow we briefly outline some of them, and outline our own developed methodological approach, as being a response to both the APIcalypse and the dominant discourse of API-dependability for research.

The approach of *auditing algorithms* was proposed by Sandvig et al. (2014) and entails different techniques to uncover the inner workings of algorithmic agents. Depending on the infrastructure and affordance of the system, the objective (input, output or system) and available resources, these techniques range from relying on APIs, use of software and hardware infrastructures to users' input, either to investigate the code or the outputs of the system. Weltevrede (2016) talks about the adoption of a *device-driven approac*h, as a way to focus on the 'the specific strategies or intents embedded in algorithms' (p. 106) and to repurpose the 'analytical affordances of the algorithmic systems/devices' (ibid.). Because algorithms are techno-epistemological devices, the analytical inquiry is dependent on the system's affordances, so on what the system allows and limits to be seen. As such it requires a combination of different types of methodological and conceptual resources to study device-captured data points. This approach shares similarities with the *reverse-engineering* one, as a diagnostic approach that allows for an observation of the relationship between the inputs and outputs, and a way to obtain 'missing knowledge' (Bucher, 2012, p. 79) and grasp a model of how the particular algorithmic system works. As a strategy to see to *what* the algorithm pays attention to, is a "process of articulating the specifications of a system through a rigorous examination drawing on domain knowledge, observation, and deduction to unearth a model of how that system works." (Diakopoulos, 2014, p. 13).

All these approaches are characterised by a move away from the quest to open the black box towards investigating *algorithms in action*, at work, in practice. It is a quest towards 'unknowing algorithms' (Bucher, 2018; Annany and Crawford, 2016), studying them as 'part of specific situations' (Bucher, 2018, p. 49) and uncovering the actor-network assemblages/configurations (Annany and Crawford, 2016). By observing the effects of the system, researchers are able to overcome the obstacles of the "black box", and to assess the 'operational principles of systems' (Bucher, 2012, p. 77) and its actual working. Additionally, investigating algorithms as an assemblage(s), to borrow Annany & Crawford's (2016) suggestion, is to look at them *as a system* and *across a/the system*.

However, this doesn't solve (all of) the difficulties of socially investigating algorithms. Algorithms are predominantly patent protected and proprietary software, with their inherent opacity stemming from the underlying machine-learning process at work. It is never a single algorithm, but always an algorithmic *system* of interconnected and interrelated algorithms (Gillespie, 2014; Bucher, 2018). In addition, these systems are in a 'perpetual beta' state (Weltevrede, 2016) with constant and continuous A/B testing, fine tuning and upgrades, making the study of algorithmic systems almost a study of a 'historical object' (Bucher, 2012). All this coupled with the research affordances (Rogers, 2013) of — and the restricted access to — the platforms' APIs and code crucially limits and impacts digital social research and complicates the task of developing and applying the appropriate methodological apparatus and tools.

Faced with these inherent characteristics, the calls for transparency of algorithmic systems, initially aiming towards total transparency, have shifted their focus significantly. Paßmann and Boersma (2017), differentiate between two notions of transparency. *Formalised transparency*, which would like to see more inside the content of the black box and 'obtain more positive knowledge' (p.140) and *practical transparency*, which does not try to open the black box, but to 'develop skills without raising the issue of openability' (ibid.). These skills should help researchers deal with the (parts of the) algorithms that we still don't have knowledge about, and probably we won't be able to have. Thus, the aim is actually to ask and investigate how to 'behave towards what remains black after all.' (ibid., p. 140). In order to find ways to work around these unknowns, the authors suggest other sources, external to the algorithms that will help turn 'unknown unknowns in to known unknowns' (p. 145), such as ethnographic data or other sources that are some kind of everyday knowledge. Our research follows the principle of practical transparency.

3 METHODOLOGICAL PATHWAYS THROUGH THE (ALGORITHMIC) SYSTEM

In his book *Design research and the new learning*, Buchanan (2001) states:

> By definition, a system is the totality of all that is contained, has been contained, and may yet be contained within it. We can never see or experience this totality. We can only experience our personal pathway through a system. (p. 12).

This corresponds with the methodological and sampling approach that we adopt in our empirical research: zooming in on a few platforms but looking at the wider system/assemblages of actors participating in the creation of an algorithmic identity of a single individual. Focusing on a research subject of one, we also expand our research to the other social and technological actors partaking in the process. In this following section we elaborate on the methodological setup while discussing the specific aspects we took into consideration and the limitations and opportunities we were faced with.

Methodological design. Our methodological approach is the result of a two-way process. First, we built our research on an assessment of the analytical affordances of the platforms in our study and of the mechanisms and tools known and available to the researchers. We tested and experimented with a variety of digital methods and tools, ranging from API-access ones to scraping ones. Through a process of going back and forth, we finalised the list of tools based on their applicability to the research questions and their particular affordances, while constantly being aware of their level of mediation (van Es et al., 2018).

Next, we experimented with the method of an *interface walkthrough* — where we mimicked and rehearsed ordinary use (*researchers as users* perspective) (Dieter and Tkacz, 2020). In that way we investigated what could be collected and used as data for research through what was available via the interface of the platforms. However, if we were to experiment with "out of the (black) box" approaches and tools, we had to think both more critically and creatively. In doing so, we "took advantage" of the newly established regulatory and transparency mechanisms and repurposed them as objects/tools for study. The platforms we queried have developed and made available (confined) gateways to transparency and explainability, as an attempt to provide more information on data collection and personalisation practices. We decided to experiment with these transparency and accountability tools and see if we could repurpose them as objects for study. Additionally, we were curious to investigate how the General Data Protection Regulation's (GDPR) Article 15, its' corresponding recitals and in particular the Data Subject Access Request mechanism (European Commission, 2016) could be used for academic research.

Approach. Faced with the above-mentioned challenges, one of the strategies created, tested and employed was to work with what is available and be creative in finding ways to do research relying on the affordances of platforms themselves and repurposing transparency and regulatory tools as objects for/of study.[1] We define these approaches as *letting the platforms*

[1] Data was collected from *Facebook*, *Google*, *Oracle*, *Quantcast* and visited webpages. The automated tools used were *AdAnalyst*, *TrackerObserver*, *PriBot* and *Privacy Score*. Data recorded included capturing of trackers on websites, *Facebook* Ads Shown, *Facebook*

speak and *making the platforms speak,* focusing on achieving *practical transparency* (Paßmann and Boersma, 2017) through investigating *algorithms in action* and studying them by observing their outputs and effects.

Sampling. The insights collected and discussed in this paper are the result of the data originating from one research subject (n=1). It is collected via different means over a period of six months[2]. Choosing the personalised, one-research-subject-only approach, allows for the observation of real user-algorithmic agents interactions, where "pre-existing profiles, browsing histories, technology fingerprints, and other organically developed profile information are used." (Bodo et al., 2018, p. 143). This real-world observation is advantageous in comparison with the use of sock puppet audits or dummy users, as it overcomes the shortcomings of 'non-adequate approximations of real-life users' (ibid.), allowing for investigation of the effects of algorithmic agents on individual users (ibid., p. 144). As such, the detailed (data) account of a personalised experience offers an overview of 'the whole spectrum of online and offline, personalised and non-personalised information flows.' (ibid., p. 145). Additionally, the insights offered by small data bear the quality of more context-aware research, granularity and depth of the data and the findings by combining various methods, complementing data and triangulating the findings (Crawford, 2013). As the method and type of data should follow the research question (Van Es et al., 2017), small data gathered using digital data analysis enables for a qualitative and contextualised investigation (Kitchin, 2014).

Focusing on algorithms in actions, around a user (a real individual, with a browsing history and data scattered around the digital space and different online and offline databases), that exhibits real-life behaviour and for whom information "in the wild" already exists, enables not just for non-lab experimentation, but also for fully taking advantage of additional non-traditional research tools, such as Data Subject Access Requests. We are aware that one of the difficulties with the auto-technographic approach is its highly individualised and personalised approach "as the observation of the interface is confined to the 'me-centric view of the researcher's own

Interests, Interactions with Advertisers (*Facebook*), Advertisers that have uploaded contact details (*Facebook*), Why am I seeing this Ad (*Facebook*), assigned interests by *Google* and reasons for assigning them. Data was collected in the period of November 27, 2018 to June 6, 2019, with different recording periods for different insights, following the browsing behaviour of one research subject. A research web browser was set prior to the start of the data collection process.

[2] Data was collected in the period from November 27, 2018 to June 6, 2019, from Brussels, Belgium. The data collection period however differs between the different tools used and the related observation. This is elaborated in more detail in the sections related to each particular tool.

account" (Weltevrede, 2016, p. 107). However, this approach both enables to manoeuvre around the "black-boxed" systems and to follow the advice by boyd and Crawford (2012) that 'the size of data should fit the research question being asked' (p. 670).

4 RESEARCHING ALGORITHMS IN ACTION

Letting the platform speak approach relies on what the platforms themselves allow to be seen and to be visible at an interface level, without the assistance or help of additional data collection tools, relying on the affordances of the presentation layer of the platforms and their *front-end*. Literary it means looking at what information platforms willingly provide and reveal via the user interface. This approach also helps uncover the platform's *politics of visibility*, i.e. what the algorithmic system itself decides to make visible and the insights it permits willingly. In addition to the focus on the interface, this approach entails use of external available sources that describe and reveal the workings of the system (Bucher, 2012a, p. 74): technical documentation, specifications, patents, media talks, but also help sections for users and advertisers. However, what we did in a novel way, and where we add to the repository of methods for research is the usage of the *transparency tools* enabled by platforms (such as Ad Settings, Data Explanations and Ad Explanations), the *privacy policies* and the *Data subject Access requests*, enabled by the GDPR. In that sense, we employed a 'multi-site technography' (Bucher, 2012, p. 73): as algorithmic systems are always assemblages and always in interaction with other actors and systems, be it technical or human, all these "sites" can be used as sources of data and insights for digital social research.

Data collection-methods wise, technography, as defined by Taina Bucher (2012) was adopted, as "a descriptive-interpretative approach to the understanding of software, rooted in a critical reading of the mechanisms and operational logic of technology." (p. 71). This is employed via observation, where the daily changes of the information provided by the platforms are observed and recorded. This approach was chosen because it allows for a granular, detailed dossier of the interaction and communication between the user and the system, it enables for insights into the actors they are *in communication* with, into what is 'the interplay between a diverse set of actors (both human and nonhuman)' (Bucher, 2012, p. 69). This is especially important for the investigation of the actor-network around the data collected and sources used for affinity profiling and algorithmic identity-building, their position within the network and in 'particular sociotechnical events' (Latour, 2005: 128 in Bucher, 2012, p. 72).

The *making the platform speak* approach, on the other hand, looks for insights not relying on what the software makes visible willingly, but by *forcing* the software to reveal itself and its inner workings. It relies on the use of automated scraping and crawling tools and tools relying on platforms' Application Programming Interfaces (APIs). In that sense it also makes visible the *politics of knowledge* of the platform, i.e. what the platform allows to be known, if one has the knowledge and tools to seek knowledge. While this can be more insightful, it is still limited. This approach aims to make the system reveal itself, in order to gain more in-depth knowledge or insights by looking not just how it produces outputs, but also to uncover things not visible at an interface level and to the human eye. In that regard, this is an analysis done at a *software* level. This approach implies that the algorithmic devices and systems will be *forced* to speak, meaning, the "analytical gaze" goes beneath the surface and what is visible and tries to uncover some inner workings of these systems.

We specifically set up a research browser through which the platforms and other actors would be able to gather as much possible information on the behaviour, actions, patterns of behaviour of the research subject and thus provide personalised search results, ads and recommendations. This enabled us to – as objectively as possible – investigate the datafication practices and the creation of algorithmic identity, while being aware of the multitude of factors affecting data collection and algorithmic outputs in the form of personalisation. In addition we were able to further investigate the assigned algorithmic identity via the outputs provided both by the used search engine (*Google)* and browser (*Chrome*) and the platforms visited during the period of the data collection phase.[3] Steps were also undertaken to allow for as much data collection and data sharing between *Facebook* and third-parties as possible, by setting up the preferences, permission and settings options[4].

[3] We set up the research browser by installing a "clean browser", deleting all the previous cookies, browsing history and preferences, and setting the preferences to allow for a maximum data collection by the platform and associated third parties: cookies were enabled, keeping record of web and app activity and location was enabled (location history, device information - info about contacts, calendars, apps, and other device data to improve users' experience across *Google* services, voice and audio activity, *YouTube* search History, *YouTube* watch history), as well as "Chrome browsing history and activity from websites and apps that use Google services" (that includes: activity from sites and apps that partner with *Google* to show ads; *Chrome* history (if Chrome Sync is turned on; app activity, including data that apps share with *Google*; Android usage & diagnostics, like battery level, how often you use your device and apps, and system errors). Ad settings were adjusted too, enabling ad personalisation, giving *Google* permission to show ads based on user's activity on *Google* services (such as Search or *YouTube*) and websites and apps that partner with *Google* to show ads. Whenever a consent by websites was asked in regard to data collection (in accordance with the GDPR), consent was given.

[4] The steps we took to set up and allow *Facebook* to maximise the data collection for the research subject were the following: changing the privacy settings and enabling data

5 INVESTIGATING DATAFICATION AND ALGORITHMIC IDENTITIES

We start our analysis by investigating datafication practices and the network of actors around the research subject. This is an important starting point, as the creation of an algorithmic identity relies on behavioural data collected via tracking elements present on both the web[5] and in apps. This step, additionally, guides the further analysis of the process of algorithmic identity creation: what data is seen as a worthy signal and what behaviour is taken as important/proxy for affinity profiling – 'grouping people according to their assumed interests rather than their personal traits' (Wachter, 2019, p. 33), based on proxies (friends, likes, groups, IP address and similar). Importantly, we are interested to see if only 'raw' data is taken as basis for inferences or there are other (hidden) mechanisms and 'cooked' data (Gitelman, 2013). The structuring of the results follows the same path: we first elaborate on our approach and findings regarding datafication and then focus on methods to investigate and assess algorithmic identity.

5.1 Investigating datafication

In order to investigate the formation of an algorithmic identity, our first step was to investigate the datafication practices surrounding a user. This provided us with insights into two interrelated aspects: the sources taken as input for the prediction outputs – the 'qualities, preferences, characteristics, intentions, needs and wants of users' (Lehtiniemi, 2016, p.4), affinities and interests — and the network of companies that collect (behavioural) data about the user (traces of user actions and interactions), as well as their dominance and variety. For this we used diverse sources of insight, collecting data on different levels (interface and software) and using a mixed method approach. We did this according to the following consecutive phases: first, using automated tools to record tracking behaviour and data collection, after which, we used privacy policies as source of information regarding data collection practices of platforms and companies. Lastly, we used transparency and regulatory tools as objects for studying datafication practices.

collection and data-sharing between Facebook family of companies and services; allowed "Ads based on data from partners"; "Ads based on your activity on Facebook Company Products that you see elsewhere"; allowed Facebook Audience Network; with the setup of the Research browser to enable third-party tracking, *Facebook* was granted access to the full browsing, off-*Facebook*, behaviour of the research subject.

[5] In this research we focus only on tracking datafication actors via web platforms.

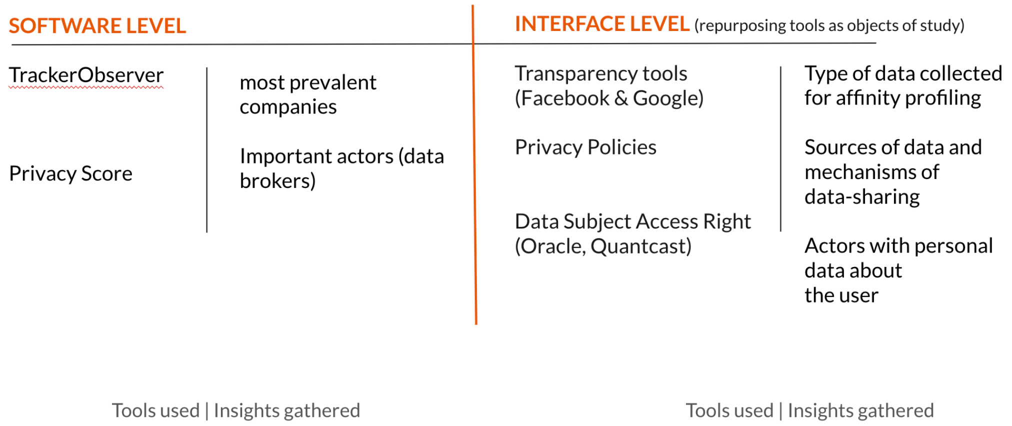

SOFTWARE LEVEL
TrackerObserver
Privacy Score
most prevalent companies
Important actors (data brokers)
INTERFACE LEVEL (repurposing tools as objects of study)
Transparency tools (Facebook & Google)
Privacy Policies
Data Subject Access Right (Oracle, Quantcast)
Type of data collected for affinity profiling
Sources of data and mechanisms of data-sharing
Actors with personal data about the user
Tools used | Insights gathered
Tools used | Insights gathered

Figure 1. Overview of the tools used for investigating datafication and insights gathered

Firstly, by using digital methods and tools, we collected information on the third-party trackers using the browser extension *TrackingObserver*[6] and the automated web scanner *Privacy Score*[7]. This was done at a software level. Both tools offer different insights in correspondence with their aim, affordances and information structure. As a result, they are suitable for different aspects and levels of analysis. Because of the ability to track every browsing behaviour around a particular user, *TrackingObserver* enables investigation of the network of third-party trackers and companies around a particular user and their unique browsing behaviour. From the data collected during a six months period[8], we obtained valuable insights into the network relations and data exchange practices of a multitude of actors. The latter was later used as a source for further investigation.

We triangulated the data obtained via the initial data collection with data available from other sources (*WhoTracksMe*[9] and *Better.fyi*[10]), providing us with several valuable insights: it enabled us to reveal the companies behind the trackers and analyse their presence, to detect the type of trackers

[6] Information about the tool is available at: https://trackingobserver.cs.washington.edu/. Last accessed January 29, 2020.

[7] Information about the tool is available at: https://privacyscore.org/. Last accessed January 29, 2020.

[8] Data was collected in the period November 27,2018 – June 4, 2019, and the analysis showed the presence of 4,691 tracking instances observed, set on 287 websites (on average 16,3 trackers per website), with 1,067 unique tracking domains.

[9] We were specifically looking at the Trackers analysis (tracking type and tracker category) and the companies indicated as owning the particular trackers. Information can be found following this link: https://whotracks.me/trackers.html.

[10] We were interested and recording the particular type of trackers, as well as the company owning the trackers. Information can be found following this link: https://better.fyi/trackers/

and their particular purpose. The analysis showed the dominance of a few companies in the network, representing the majority of trackers on the visited websites (Figure 2).

However, we also observed a long tail of many different actors (a large number of trackers with low websites frequency) that captured data about the user's behaviour, supporting similar findings by Binns et al. (2018b). Categorising the detected trackers based on a taxonomy, we discovered a presence of a vast and well-developed network of *ad networks*, counting for more than half (57.23%) of the detected unique trackers. These findings are important for several reasons. The detected long tail is worrying as it indicates that a great number of companies get some and partial data from the research subject and users in general. This is even more of a cause for concern as the profiling-oriented businesses, being faced with lack of informational awareness and with ‘information gaps’ (Crain, 2018, p. 91), need to infer data and predict behaviours using analytics and modelling to fill that gap. If these sources are ‘data poor’, the inferences and algorithmic identities (poorly) built on them will be inevitably inaccurate, affecting further the automated decision-making processes.

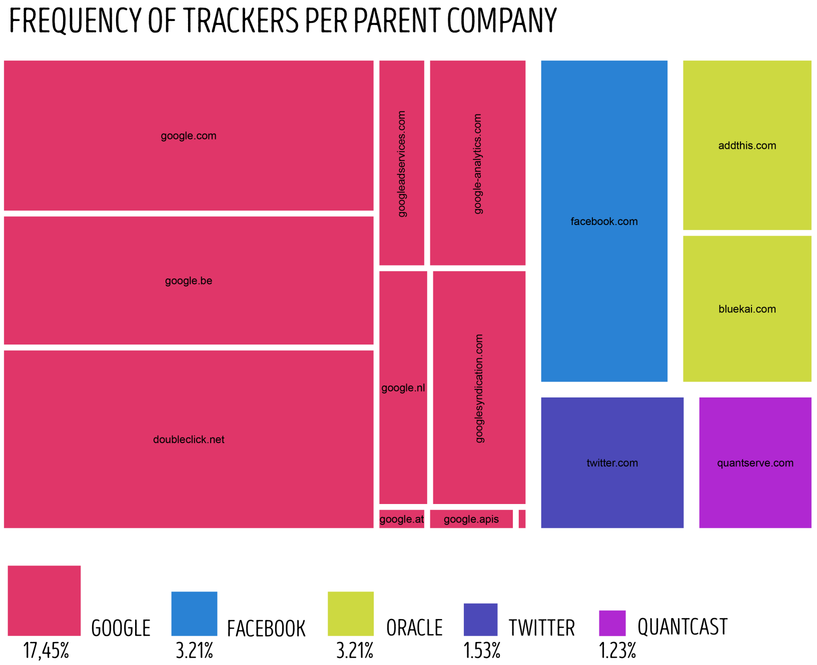


Figure 2. The most prevalent trackers per company in the research dataset as captured by TrackingObserver, data triangulated with WhoTracksMe and Better.fyi[11]

[11] For the analysis and the visualisation, we focused only on the most prevalent trackers by company, in order to detect the most dominant ones collecting behavioural data about the research subject. That is the reason why the percentages don’t sum up to 100% and the long tail is not fully visualised.

Privacy Score[12] provided us with different insights. Aiming to investigate which are the websites that capture the user's habits, what kind of trackers are present and for what purposes, we scanned the top 10 most visited websites by the research subject. As detected with the *TrackerObserver*, if we look at the presence of company trackers in the sampled websites, here we also encounter a well-developed network, dominated by *Google* and distantly followed by *Amazon, Oracle, Facebook, Conde Nast and Quantcast* (Figure 3). The analysis further shows that most of the trackers set by third parties are via cookies (73,41%) and for the purpose of advertising (83,23%) (Figure 4). Cross-referencing data collected via *Privacy Score* with data from *Better.fyi* and *WhoTracksMe* enabled us to detect the purpose for tracking and the tracking type detected (Figure 4). This additionally confirms that most of the surveillance done online is for the purpose of accumulating data for online behavioural advertising, referring to personalised and targeted advertising based on prediction of interests and affinities profiling.

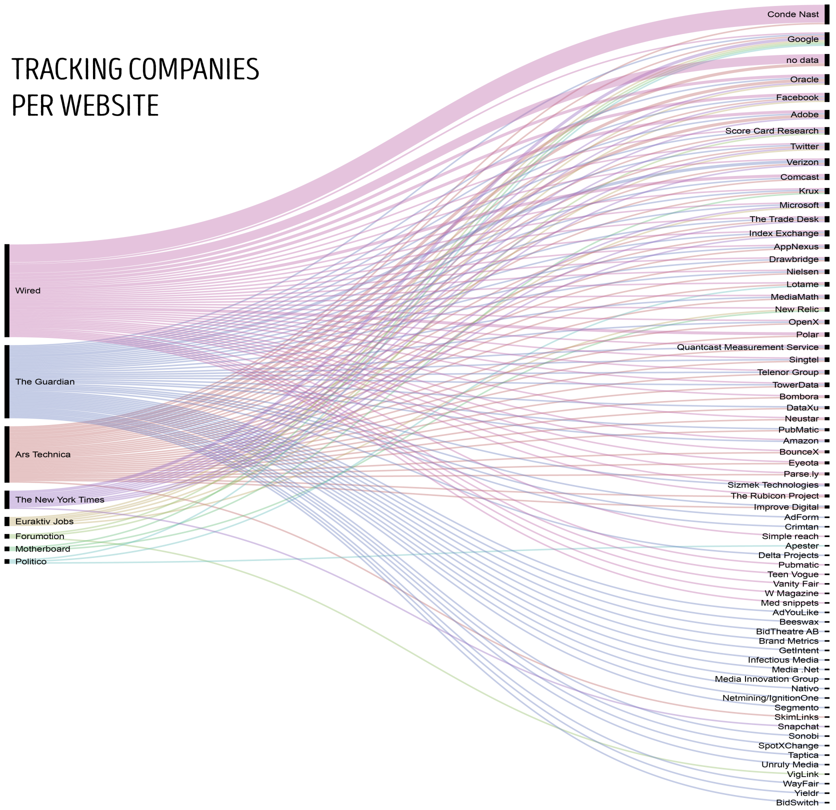


Figure 3. Frequency of third-party trackers per website. Data source: Privacy Score

[12] Data collected with Privacy Score was done one-time only, as the presence of trackers is tied with the website, not the research subject. The data collection was done on February 21, 2019 and it reflects the state of the particular website at that particular data. Data collected showed that the number of third party embeds (third parties that provide services to the first party) is 575 for only ten websites, set by 328 unique companies, and the number of third-party calls is 172.

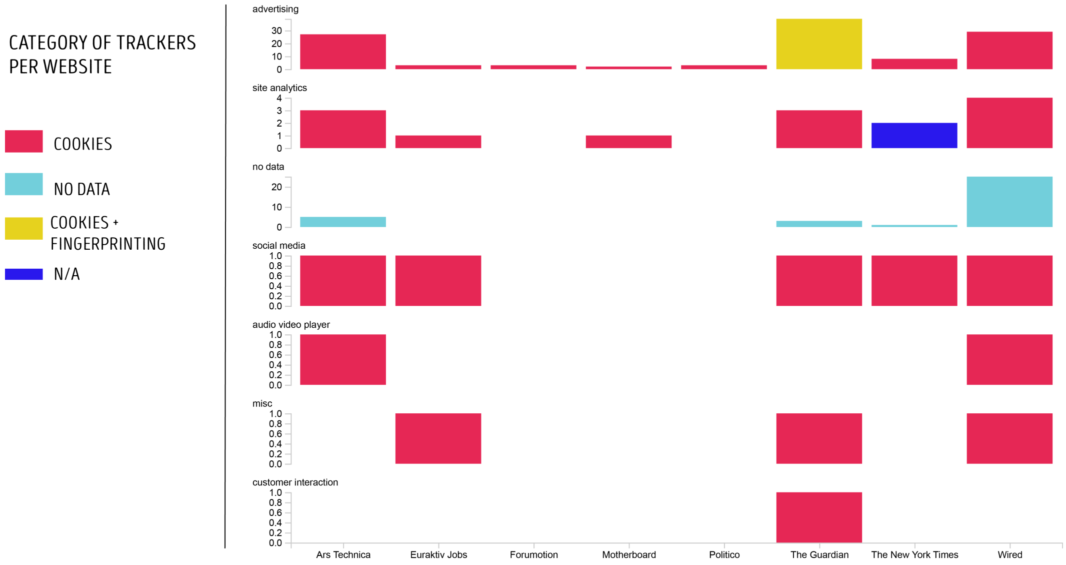


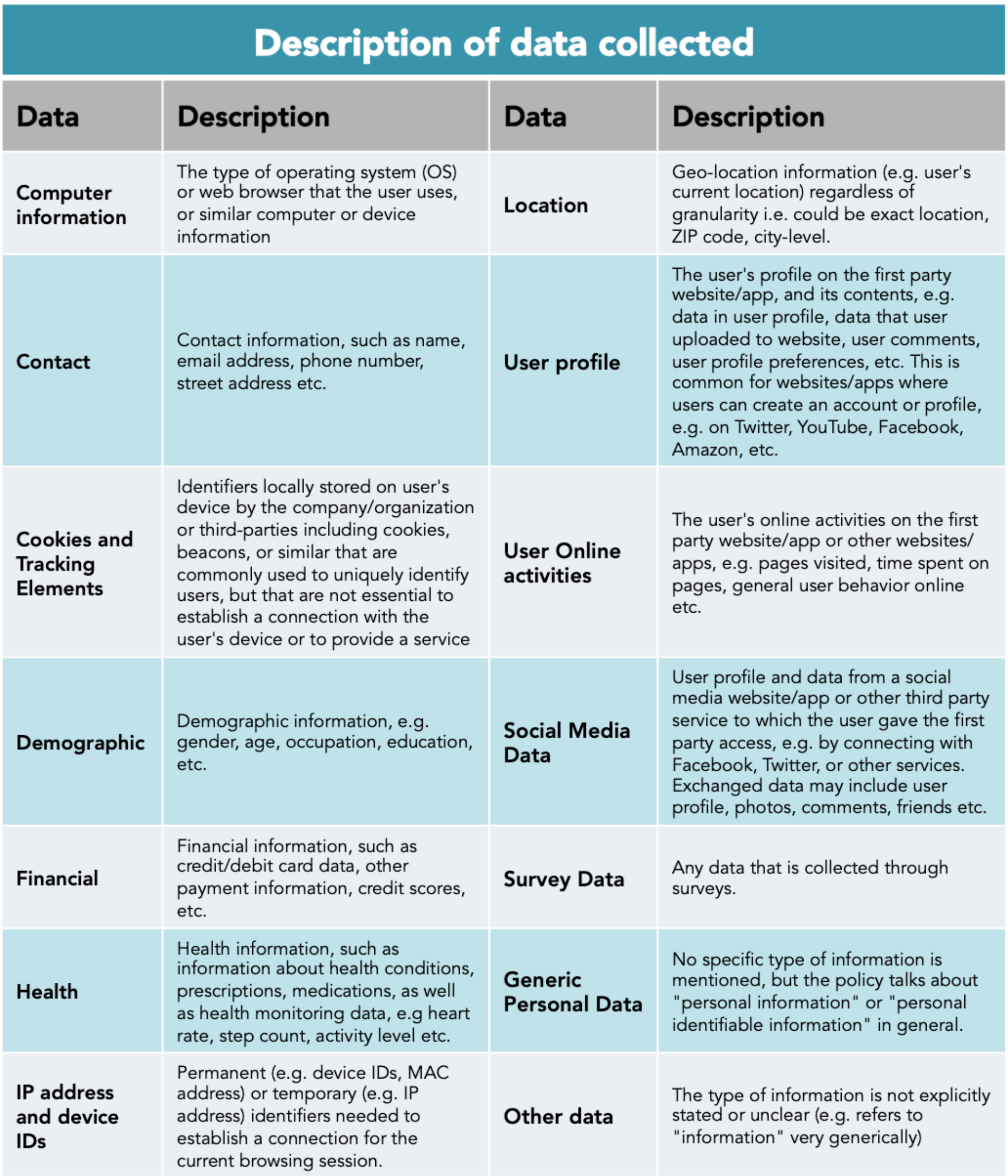
Figure 4. Categories of trackers per website. Data source: Privacy Score, data triangulated with WhoTracksMe and Better.fyi

The insights collected from these two tools guided the subsequent research steps. It was expected that *Google* and *Facebook* would be the most prominent companies. However, observing the not-insignificant presence of data brokers such as *Oracle* and *Quantcast* motivated the further investigation about the data these companies hold about the research subject and the algorithmic identity they assigned. Data brokers are important actors since they

> are businesses whose revenue model revolves around aggregating information about individuals from a variety of public and private sources […] who sell access to the collected data to third parties, including advertisers, marketers, and political campaigns. (Venkatadri et al., 2018, p.1).

We investigated their role and the data they have by looking at what is detectable regarding datafication practices by different actors at an interface level. To do so, we experimented with data from less-traditional sources: the privacy policies of the most dominant tracking companies we detected in the previous step, the transparency tools made available by the actors themselves and the regulatory tools — the Data Access Request mechanisms enabled by Article 15 of the GDPR.

We started with the *privacy policies* as investigation tools. We sampled the following platforms — *Google* and *Facebook* — and two data brokers — *Oracle* and *Quantcast* — detected previously. To get better initial structured overview, we used the machine learning tool, *PriBot*[13], in order to collect data on (1) what kind of data is being collected about the users and (2) the reasons for data collection. Although privacy policies can be information-rich sources, we decided to narrow our analysis to these two aspects only, as they are the most relevant for our research question.



Description of data collected

Data	Description	Data	Description
Computer information	The type of operating system (OS) or web browser that the user uses, or similar computer or device information	Location	Geo-location information (e.g. user's current location) regardless of granularity i.e. could be exact location, ZIP code, city-level.
Contact	Contact information, such as name, email address, phone number, street address etc.	User profile	The user's profile on the first party website/app, and its contents, e.g. data in user profile, data that user uploaded to website, user comments, user profile preferences, etc. This is common for websites/apps where users can create an account or profile, e.g. on Twitter, YouTube, Facebook, Amazon, etc.
Cookies and Tracking Elements	Identifiers locally stored on user's device by the company/organization or third-parties including cookies, beacons, or similar that are commonly used to uniquely identify users, but that are not essential to establish a connection with the user's device or to provide a service	User Online activities	The user's online activities on the first party website/app or other websites/ apps, e.g. pages visited, time spent on pages, general user behavior online etc.
Demographic	Demographic information, e.g. gender, age, occupation, education, etc.	Social Media Data	User profile and data from a social media website/app or other third party service to which the user gave the first party access, e.g. by connecting with Facebook, Twitter, or other services. Exchanged data may include user profile, photos, comments, friends etc.
Financial	Financial information, such as credit/debit card data, other payment information, credit scores, etc.	Survey Data	Any data that is collected through surveys.
Health	Health information, such as information about health conditions, prescriptions, medications, as well as health monitoring data, e.g heart rate, step count, activity level etc.	Generic Personal Data	No specific type of information is mentioned, but the policy talks about "personal information" or "personal identifiable information" in general.
IP address and device IDs	Permanent (e.g. device IDs, MAC address) or temporary (e.g. IP address) identifiers needed to establish a connection for the current browsing session.	Other data	The type of information is not explicitly stated or unclear (e.g. refers to "information" very generically)

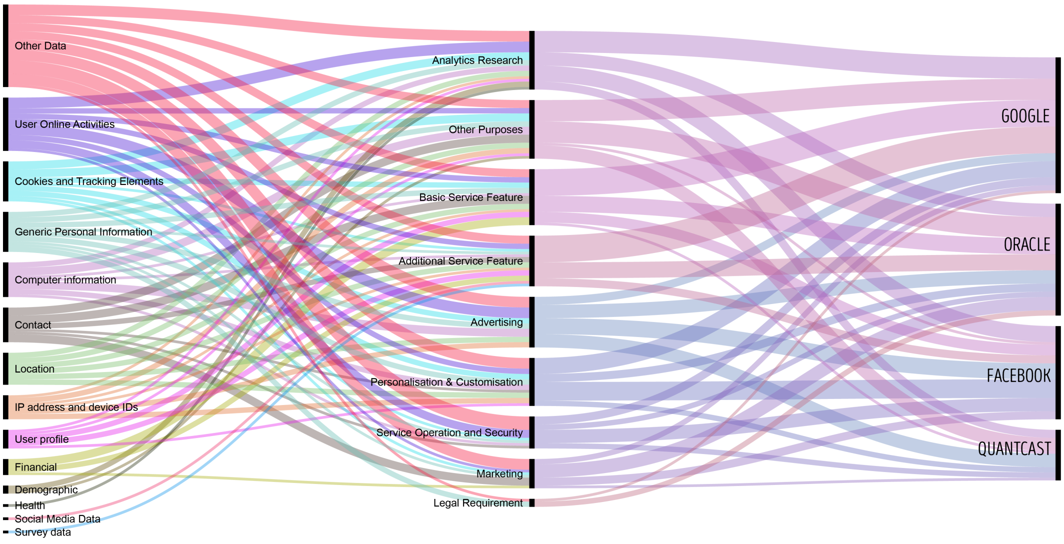
Figure 5. Overview of the type of data and specification of data types in the sampled privacy policies (information source: PriBot)

By analysing and comparing the information we obtained from the *PriBot* tool, a list of all possible data types that could be collected by these actors,

[13] *PriBot* is an AI-powered tool for automated analysis of privacy policies https://pribot.org/polisis

the above table was created (Figure 5), listing all the data and their definitions, that were/could be captured for the research subject and further (re)used, and which are of particular interest to the sampled companies. This reveals a dominant 'collect all' approach, where the (legal) principle of data minimization is not respected and a lot of data that is not necessary for establishing a connection or providing a service is captured.

We additionally analysed and cross-referenced the findings for each of the actors (Figure 6), thus being able to discover the relations between the types of data collected by each of the sampled policies, the stated reason for collection and the actors that collect each type. The analysis shows that the most under-defined category — "Other data" is the most frequently captured data, although it was not explicitly stated in any of the policies what kind of data that is, leaving many open doors for misuse and abuse. Looking at the column with particular actors, it is noticeable that apart from *Google* (not unexpected), *Oracle* is actually the actor that closely follows *Google* for potential capture of a number of various data types. Figure 6 shows how "messy" data collection is, and how different types of data can be used for various purposes. We can detect, for example, that *Personalisation & Customisation* is a reason for data collection for all sampled companies, and the following types of data are used for that purpose: user profile, IP address and device IDs, location, contact information, computer information, generic personal information, cookies and tracking elements, user online activities and other data. For *Marketing purposes*, companies use financial data, contact information, generic personal information, cookies and tracking elements, user online activities and other data.



TYPES OF INFO COLLECTED AND REASON FOR COLLECTION PER COMPANY

Figure 6. Diagram of data collected and stated reasons for collection across companies (information source: PriBot)

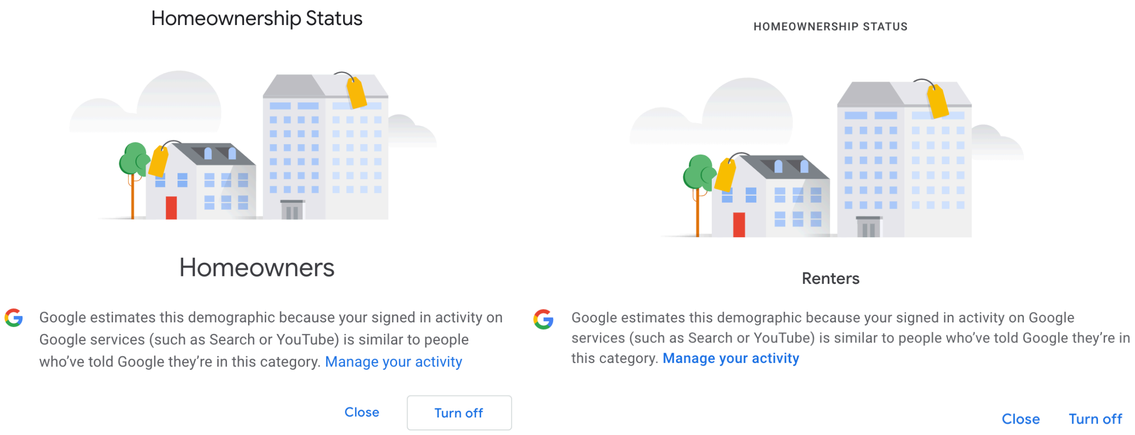
What we call *transparency tools* are designed by the platforms with a specific purpose in mind: to increase transparency and accountability towards users and regulators (Facebook Newsroom, 2020; Google Blog, 2018). However, here we are repurposing them as *objects for study* in order to investigate datafication practices and sources. For this particular case, we looked into *Google* and *Facebook's* data explanation and ads explanation mechanisms.

Google's Ad preference page[14], for example, shows the inferred interests about each particular user, briefly elaborating on the logic and process behind it. This allows us to investigate where the (behavioural) data originates from. Having this information, we can see how data is captured and transferred and thus get insights into the datafication and data sharing network. Following and recording the data a few times a week over a period of two and a half months[15], during which we collected 183 distinct interests assigned to the research subject, our research showed that *Google* estimates the interests based on using and/or combining data from: 1. activity on *Google* services/products; 2. activity on *Google* combined with activity on other websites and apps; 3. activity on non-*Google* (outside *Google*) services and 4. Visiting an advertiser's website/app.[16] This also gives insights into the structuring of information and the degree of (non)disclosure by the platform itself, impacting the degree and scope of possible research insights. However, as these systems are highly volatile, at the time of writing this article and checking explanations again, it was noticed that *Google* added one more insight source — "similarity to other users". As an example, for the categorisation "Homeownership Status" *Google* categorises the research subject as "Renter" based on "Google estimates this demographic because your signed in activity on Google services (such as Search or YouTube) is similar to people who've told Google they're in this category". Additionally, three months before, the research subject was categorised as "Homeowner", based on the same sources (see Figure 7).

[14] It can be accessed at the following link: https://adssettings.google.com/authenticated

[15] This data was collected in the period of March 2, 2019 to May 17, 2019.

[16] The explanations provided by Google for each of the sources are the following: 1. Google services/products - "Google estimates this interest, based on your activity on Google services (such as Search or YouTube) while you were signed in"; 2. Google and other providers - "Google estimates this interest, based on your signed-in activity on Google services (such as Search or YouTube), as well as on your signed-in activity on non-Google websites and apps"; 3. non-Google (outside Google) - "Google estimates this interest, based on your activity on non-Google websites and apps while you were signed in"; and 4. Visited advertiser - "This advertiser shows you ads based on: Your visit to the advertiser's website/app".



Google estimates this demographic because your signed in activity on Google services (such as Search or YouTube) is similar to people who've told Google they're in this category. Manage your activity

Google estimates this demographic because your signed in activity on Google services (such as Search or YouTube) is similar to people who've told Google they're in this category. Manage your activity

Figure 7. Screenshots from the research subject's Google's Ad Settings page. The one of the left dates from April 27, 2020; the one on the right is from July 21, 2020

Facebook offers more transparency mechanisms, of which we used the *data explanations*[17] and *ad explanations*[18]. We used these two tools to collect information on the sources of data, the type of data (whether or not personal data) and the actors in the datafication network, as well as — equally important — the mechanisms and sharing practices between the actors in the network.

The insights provided show that *Facebook* datafies users both on- and off-platform, of which the latter one is the prevalent one. Using additional sources of insights about the workings of the platform's tracking system, such as guidelines offered to advertisers by *Facebook* itself, shows that this is data originating from the websites integrating the *Facebook Pixel* tracking technology, and is handed to the platform by clients (websites/app) that integrate it. Clients uploading a contact list to *Facebook* is another source of data feeding the platform. These two sources (*Pixel* and *List*) contain personal data and they constitute 68.25% of the off-platform data ending up at *Facebook*. The only data originating from on-platform behaviour is the data gathered by tracking the ads shown on *Facebook's* Newsfeed that were clicked. Recording the data available via the "advertisers who use contact list added to *Facebook*" tab, shows that very high percent (75%) of companies

[17] Data explanations provide the user with a list of attributes Facebook has inferred about them, how they were inferred and what information is used to target them with advertisements (see Andreou et al., 2018 for more detailed explanation of the mechanisms). The data explanations are accessible via an Ad Preferences Page (https://www.facebook.com/ads/settings) and they provide information structured in the following way: Your interests, Advertisers and Businesses, Your information and Ad Settings.

[18] Ad explanations provide the user with an information/explanation why a particular ad was served. They are accessible via the "Why am I seeing this?" button above every ad served on the user's Newsfeed.

listed collected personal data from other sources, not the user itself, without the user's explicit consent or information about the source provided. This potentially points to the well-developed network of actors in the (personal) data sharing network.

Repurposing the *Ad explanation* tool by *Facebook*, particularly the "Why am I seeing this ad" option, we were able to collect information on the data sources used for personalised ad targeting.[19] We did this on both levels (interface and software), using both observation for recording the data from the interface, and the automated tool *AdAnalyst*, to collect data at a software level. Following an analysis of the explanations provided, we were able to uncover the relations between the sources of data used, the types of data used, the analytical processes at play and the particular reasons for personalised ad targeting, shown in the figure below (Figure 8). For example, if the targeting is based on a *particular interest*, behavioural data will be used to make that inference. This data could be originating either from *Facebook* (by tracking the activity of the user), and advertisers and/or data brokers, using inferential and prediction analytics. The latter analytics methods are used to infer user preferences, attributes and opinions and predict behaviour (Wachter and Mittelstadt, 2018, p.4). Reading, structuring and coding the information collected and recorded, provided us with additional insights: apart from insights into the processes behind the ad-targeting analytics and the inputs/outputs relations, it also revealed that the sources of data could originate both on- and off-platform, they can be volunteered (by the user), obtained (via partners, data brokers and advertisers) or captured by *Facebook*. Different types of data are taken as signals for affinities/interests. This ranges from location and age, to languages spoken, activities and social neighbourhood, or tracking the social network of/for relations between individuals/users and taking this as a data signal for further affinity profiling and commodification for ads targeting. This 'data inference process' (Andreou et al., 2018, p.3) is important because it allows the advertising platform to infer users' preferences and attributes, later used for affinity profiling and building algorithmic identity, further used as a basis for commodification (targeted advertising).

[19] Such as: liked advertising page, visited advertiser's website or app, friends liked a page, age/gender/location, activity on Facebook's family of apps & service, particular interest etc.

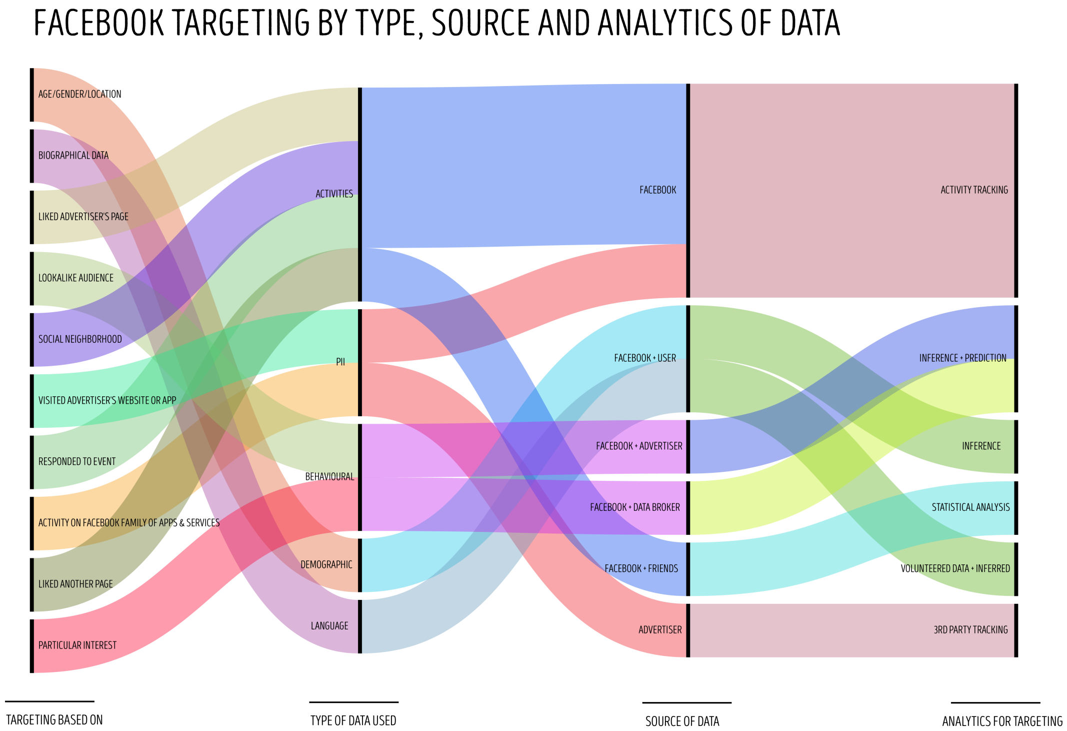


Figure 8. Alluvial diagram of sources of data and inferences for Facebook

The last strategy we used for uncovering and investigating the data sources, actors and mechanisms for inferential analytics, prediction and building algorithmic identity was repurposing the *Data Subject Access Rights* mechanisms as an object for study. Article 15 of the GDPR, in force since May 2018, enables data subjects to request and obtain access to any personal data being held and processed by a data controller. Executed in correct manner, it should give information on the purposes of the processing, the categories of personal data concerned, the recipients or categories of recipients to whom the personal data have been or will be disclosed, and if automated decision-making (including profiling) is present. The latter entails providing meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject (European Commission, 2016). Repurposed for academic research, Data Subject Access Requests (DSAR) would give information on the sources of data (categories of personal data concerned), the network of actors with access to the data and the algorithmic identity/assigned affinities by the controller.

Six DSARs were filed, of which only one response (by *Oracle)* was entirely suitable for analysis.[20] The data obtained from *Quantcast*, although

[20] Requests were sent to *Bumble*, *Oracle*, *Criteo*, *Quantcast*, *Facebook* and *Acxiom*. Only *Oracle* provided data that can be used for the purposes of the research. The file obtained by *Quantcast* was "unreadable" in terms that it contained only a few unique rows, duplicated tens of thousands of times (96,659 data entries in total). *Criteo* was asking for additional identification checks, and because of time constraints it was decided not to follow through.

incomplete, enabled for some crucial observations. The first observation pertains to the well-established and wide network of data sharing and the exchange system between data brokers. *Oracle* relies on six other data brokers to collect data and infer affinities and interests (these data brokers are *Eyeota, OnAudience, Lotame, Bombora, AuDigent, Affinity Answers*). This complicates the quest of tracking where data originates and where its' final destination is, making it difficult to later contest or rectify the data in question. The second observation concerns the risk of inaccurate inferences: if one data broker makes inaccurate inference, this information is further shared across the ecosystem. Closely inspecting the data provided by *Oracle*, it could be observed that some of the inaccurate data *Oracle* holds originates from *Eyota*, that obtained them from *Bombora*. The reliance on other partners and data brokers is also indicated in the data obtained from *Quantcast*, in their "Audience Grid" data file, which points to a largely adopted practice. This might have serious consequences for the data subject resulting in not just their erroneous profiling, but also (potentially) in access to services and opportunities.

The "unsuccessful" DSARs also demonstrate that the access to personal data held by online platforms is more often than not a complex and uncertain process. Because of the different interpretations of the DSAR procedure and the GDPR in general by companies, there are apparently substantial differences about what data is considered personal and thus eligible to be provided by the data controllers.[21] Sometimes the data controllers have long and extensive procedures (like *Criteo*) or they try to bypass meaningful information by directing users towards other available data (*Facebook*). Even when successful, the data obtained might not be readable (as in the *Quantcast* case), the file might be incomplete, and the logic behind the presented and provides data and information might not be available or accessible for the user.

5.2 Investigating algorithmic identity

Next we investigated the workings of the algorithmic systems of a web platform (*Google*), social media (*Facebook*) and one data broker (*Oracle*). We took the inferences as proxies, or represents, for investigating the assigned

Facebook provided data, but with no additional meaningful information, and the data corresponds with the one provided on their platform via the "Download your information" tool. *Acxiom* provided an answer stating that no data is collected from individuals residing in Belgium.

[21] In an attempt to obtain data from the dating app *Bumble*, the platform representatives stated that they can only provide a registration date, IP addresses and profile photos (source: personal correspondence).

algorithmic identity. We decided for sampling these two platforms and the data broker based on the results from the datafication phase of the research, where most trackers were originating from these three actors (and as such have most data on the research subject), and on their affordances for research.

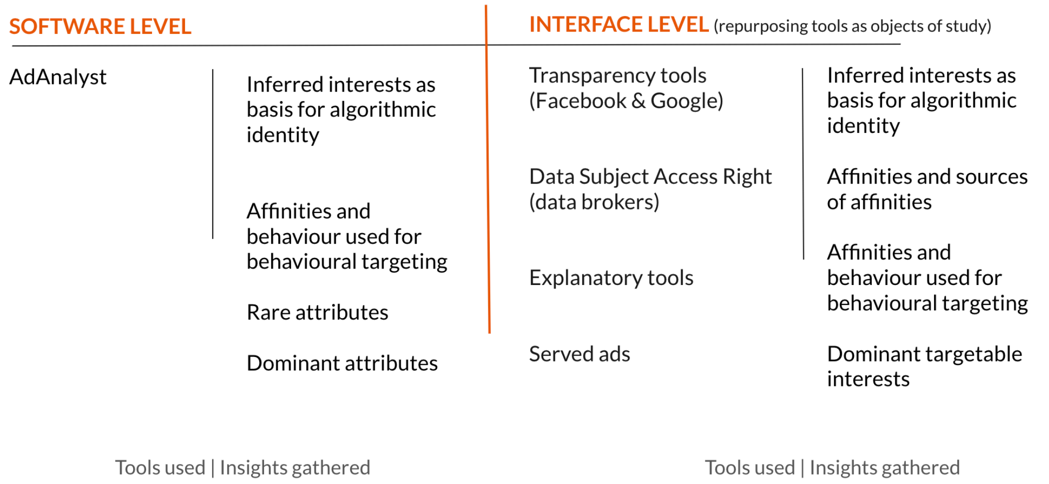

SOFTWARE LEVEL
INTERFACE LEVEL (repurposing tools as objects of study)
AdAnalyst
Inferred interests as basis for algorithmic identity
Affinities and behaviour used for behavioural targeting
Rare attributes
Dominant attributes
Transparency tools (Facebook & Google)
Data Subject Access Right (data brokers)
Explanatory tools
Served ads
Inferred interests as basis for algorithmic identity
Affinities and sources of affinities
Affinities and behaviour used for behavioural targeting
Dominant targetable interests
Tools used | Insights gathered
Tools used | Insights gathered

Figure 9. Overview of the tools used for algorithmic identity and insights gathered

The three different data controllers are investigated in order to assess the assigned algorithmic identity and test the possibility for research using the inferred interests as proxies. As Figure 9 shows, we employed a variety of methods and tools, at a different level, to analyse various aspects of the inferential analytics at play and their outputs.

Google's Ad settings tool[22] was observed in frequent intervals for two and a half months and it was used to record the assigned interests. Based on the 184 observed interests, and triangulated with a list of categories (Brave, 2019) indicating the particular category an interest belongs to, enhanced with a close reading of the categories, we were able to get an overview of the most dominant categories the research subject was categorised in (Figure 10). The daily recording of the interactions and assigned interests show that these are often immediate outputs of simple browsing behaviour, but also that they are unstable and disappearing – thus no historical database of inferred interests is available (for research or personal insights). Some of the interests disappear on a daily basis and some remain longer periods of time, or during the entire period of data collection. This is significant from a point of view of reliability of collected data: researchers must be aware of the instability of the data and the potential inability of collecting what is available. This underlines the

[22] The information available by the platform was monitored, collected and recorded in the time period of March 2, 2019 – May 17, 2019, and 183 interests assigned were observed.

dependence on and significance of the information structuring and information visibility, which can be seen as politics both of visibility and knowledge, controlled by the platforms themselves. Andreou et al. (2018) point to the same characteristic of *Facebook's* transparency tools, referring to it as *snapshot/temporal completeness*.

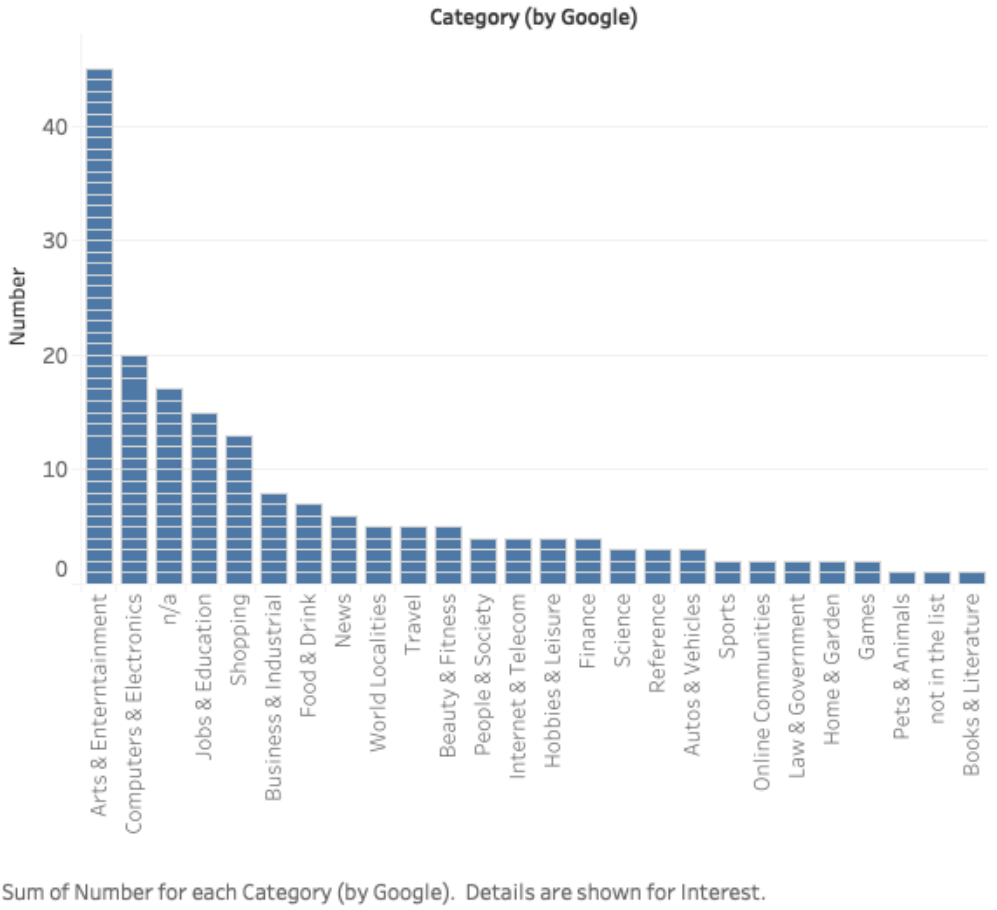


Figure 10. Frequency of categories of interests as assigned to the research subject by Google

Reading the assigned interests as *text*, we were able to construct an overview of the assigned algorithmic identity by *Google* (Figure 11). The use of the auto-technographic approach, as well as the fact that we are relying on and working with data from a real individual, enables us to test the assumptions made by the algorithms and assess its truthfulness. In our case, the assigned algorithmically constructed identity is in a sharp discrepancy with the research subject's sense of real identity and does not represent their actual life conditions (financial, familial, or employment). Similar are the findings from the data collected from *Oracle*, with the important difference here that online data brokers often lack information on basic demographic data and thus have to infer it via browsing behaviour in order to fill the 'information gap' (Crain, 2018), unlike platforms like *Google* and *Facebook* that rely on both volunteered data (by users) and have more access to daily behaviour of users. However, we must be aware as researchers that an important aspect of reading and interpreting the data is concealed by the platforms: there is a lack of information on how these attributes are assigned, and what is the inferential analytics process. This potentially affects the comprehensiveness of the data collected by the researcher and consequently — the analysis itself.

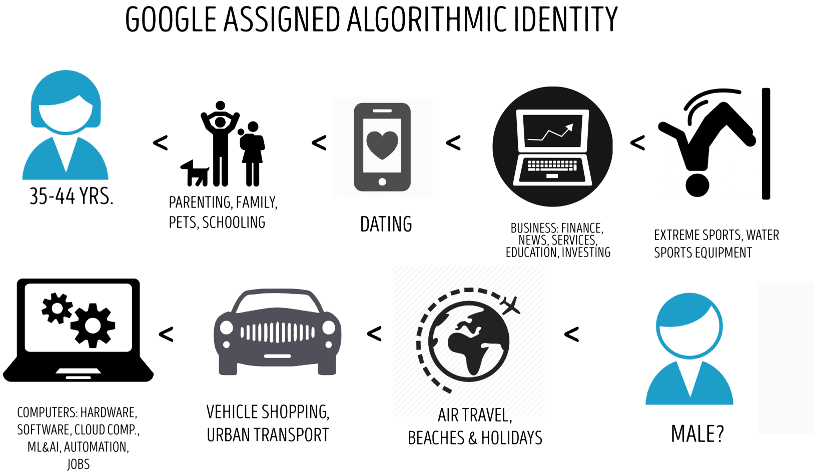

GOOGLE ASSIGNED ALGORITHMIC IDENTITY
35-44 YRS.
<
PARENTING, FAMILY, PETS, SCHOOLING
<
DATING
<
BUSINESS: FINANCE, NEWS, SERVICES, EDUCATION, INVESTING
<
EXTREME SPORTS, WATER SPORTS EQUIPMENT
COMPUTERS: HARDWARE, SOFTWARE, CLOUD COMP., ML&AI, AUTOMATION, JOBS
<
VEHICLE SHOPPING, URBAN TRANSPORT
<
AIR TRAVEL, BEACHES & HOLIDAYS
<
MALE?

Figure 11. A close-reading illustration of an algorithmic identity as assigned by Google

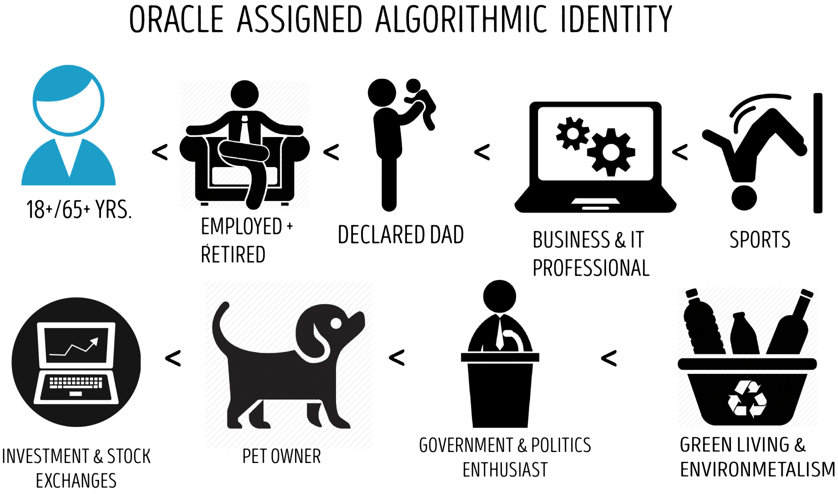

ORACLE ASSIGNED ALGORITHMIC IDENTITY
18+/65+ YRS.
<
EMPLOYED + RETIRED
<
DECLARED DAD
<
BUSINESS & IT PROFESSIONAL
<
SPORTS
INVESTMENT & STOCK EXCHANGES
<
PET OWNER
<
GOVERNMENT & POLITICS ENTHUSIAST
<
GREEN LIVING & ENVIRONMETALISM

Figure 12. A close-reading illustration of an algorithmic identity as assigned by Oracle

When it comes to the possibility to investigate algorithmic identity as assigned by *Facebook*, by using the very affordances of the platform itself, we were able to draw an overview of the assigned general affinity towards certain categories, via few available "points". We used as data source the *data explanations* (revealing the reason for assigning the interests[23]) and the *ad explanation* feature, both at an interface level (via observation and recording of data) and at software level (*AdAnalyst* tool).

The data collected via the data explanation feature, gives not only insights into the dominant assigned interests by category (Figure 13), but also points to the very specific categorisation practices *Facebook* uses for profiling and targeting. Closely reading the list of interests, it becomes visible that *Facebook* is constructing very narrow categories (e.g. *headphones; old style and new style dates; Conversion (gridiron football); Right-to-work law*; particular movies/songs etc.) that might enable a very specific targeting, and also, that many of them simply do not make sense (e.g. *non-resident Indian and person of Indian origin*; *hydrogen*) or can be regarded as potentially sensitive information (*Gay Pride, LGBT community*).

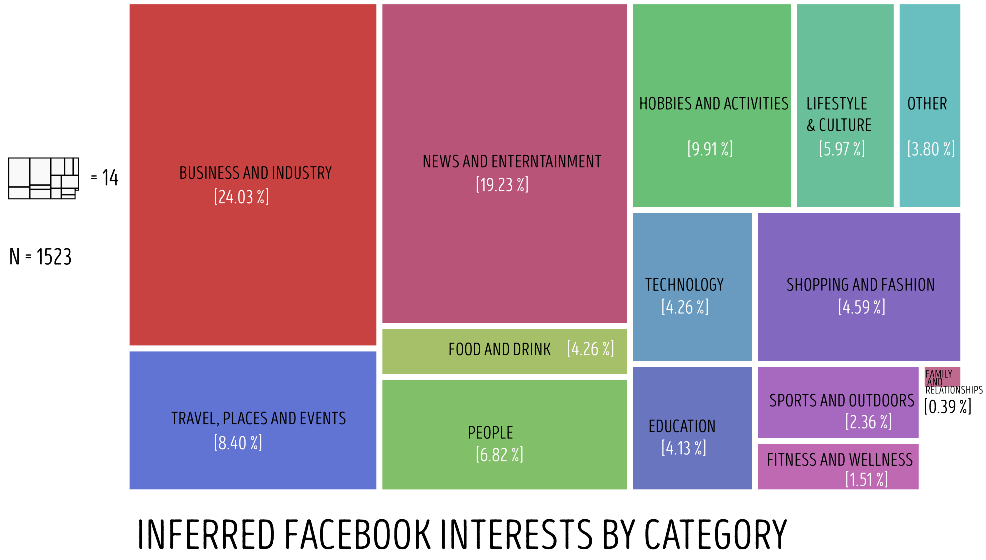


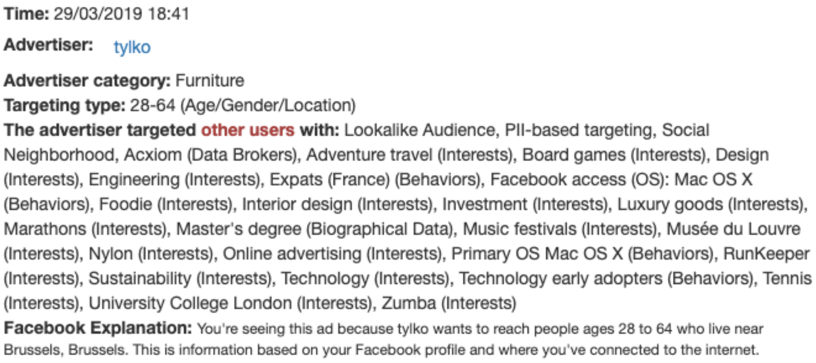
Figure 13. Dominance of interests assigned by Facebook, per category

Although the matching process of a user being served a particular ad is complex due to the fact that the outcome doesn't depend only on the advertising platform and its matching algorithm, but also on the very event-specific factors[24], the explanatory tool "Why am I seeing this ad?", when

[23] At the time of the data collection and analysis of the *Facebook* data (11/27/2018 - 04/16/2019), *Facebook* was providing three very generic explanations why an interest/affinity was inferred, but no further information: "you clicked an ad related with the interest" (64.37%), "you liked a page related with the interest" (31.80%) or "installed an app" (0.5% of the entries). Additionally, it added "liked their page or post" (3.30%) - data recorded in May 2019 showed that *Facebook* made changes to their "assigned interests" explanation, adding one more "reason" to the previous three.

[24] Such as the competing advertisers at the particular moment when an ad is about to be served, their specific requests/objectives set by the advertisers and the characteristics of the available users on the platform, in a particular moment of time (Andreou et al., 2018, p. 3).

repurposed as an object for study, can provide significant information regarding the particular behaviour, activities and interests of the research subject used for automated behavioural targeting. Combining the insights collected manually via the interface with the data collected automatically via *AdAnalyst* at a software level, provided significant insights. The first finding is related with the type of data that algorithmic systems consider as an important particular aspect of the research subject's algorithmic identity to be later taken into account for personalised behavioural ad targeting.[25] The second one relates to the affordances of the different research methods and tools, and the different insights, depth and scope of insight that they enable. *AdAnalyst* offers different insights as it has access to more parameters at a software level, not accessible via the interface. Such is the distinction between the general ad explanation served to the research subject (as a user) and what is indicated as a reason the particular user to be targeted. Additionally, insights can be obtained about the targeting parameters set by the advertisers. As Figure 14 shows, what the research subject has been targeted based on (e.g. bicycle as interest), might be just one of the campaign targets set by the advertiser. These can sometimes be different, and in that sense *AdAnalyst* provides more in-depth insights than available if looking only at an interface level.



Time: 29/03/2019 18:41

Advertiser: tylko

Advertiser category: Furniture

Targeting type: 28-64 (Age/Gender/Location)

The advertiser targeted other users with: Lookalike Audience, PII-based targeting, Social Neighborhood, Acxiom (Data Brokers), Adventure travel (Interests), Board games (Interests), Design (Interests), Engineering (Interests), Expats (France) (Behaviors), Facebook access (OS): Mac OS X (Behaviors), Foodie (Interests), Interior design (Interests), Investment (Interests), Luxury goods (Interests), Marathons (Interests), Master's degree (Biographical Data), Music festivals (Interests), Musée du Louvre (Interests), Nylon (Interests), Online advertising (Interests), Primary OS Mac OS X (Behaviors), RunKeeper (Interests), Sustainability (Interests), Technology (Interests), Technology early adopters (Behaviors), Tennis (Interests), University College London (Interests), Zumba (Interests)

Facebook Explanation: You're seeing this ad because tylko wants to reach people ages 28 to 64 who live near Brussels, Brussels. This is information based on your Facebook profile and where you've connected to the internet.

Figure 14. Screenshot of AdAnalyst's interface

The screenshot above is interesting for analysis because, via the section "The advertisers targeted other users with", it provides valuable insights into the parameters *Facebook* uses for targeting. We can observe that apart from the well-known Lookalike audience, Personally Identifiable Information (PII)[26], Social Neighbourhood and similar, it also targets users

[25] For example, the research subject has liked a page, has or was at a particular location, belongs within a particular age group, etc.

[26] Personally Identifiable Information (PII) is considered any data that can be used to identify a specific individual, such as name, email, phone number, IP address, location address, online identifier, biometric records and similar. For more detailed definition, see GDPR Art. 4 (1).

based on data from data brokers, based on behaviours (e.g. expats in France), operating system and version (based on where *Facebook* was accessed from) and biographical data (Master's degree).

Another avenue to investigate and assess an assigned algorithmic identity is to repurpose the particular ads served to the user, more specifically the textual part of each of the ads. As the purpose of the ads is to nudge users to take particular action, ads are served targeting specific interests of particular users, with the aim to steer actions or behaviour. In that sense, ads could uncover the assigned affinities and, at an aggregate level, the algorithmic identity. Thus, a semantic analysis of 1,553 served ads, collected both manually (interface level) and using *AdAnalyst* (software level), was done. Only unique ads were taken into consideration. The tool *CorText* (Munk, 2019) was used to detect the semantic clusters forming from the corpus of served ads. The frequency of the semantic co-occurrence can be read as a signal of attributes the user is more targetable for, or most prone to take actions for. It can also be seen as enabling an insight into how a particular user is seen by the algorithms, given that the most dominant reasons for targeting are being part of a lookalike audiences and because of specific user interests.

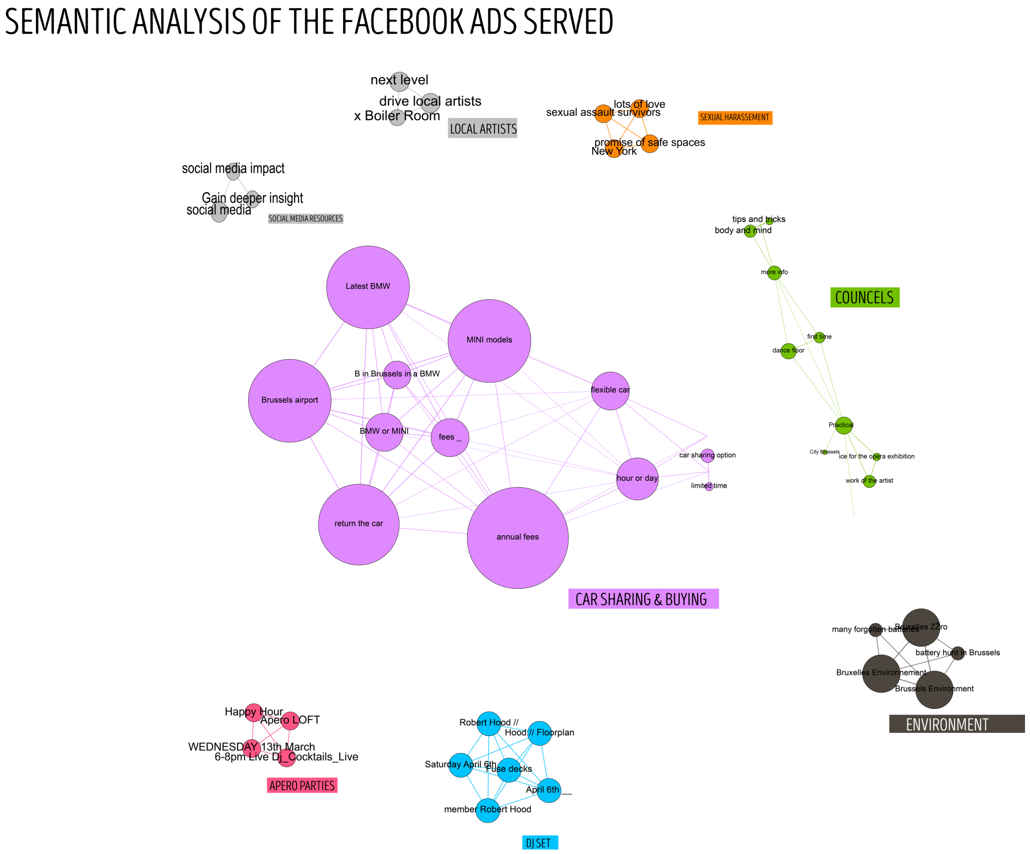


Figure 15. Network mapping of semantic clusters from served ads on Facebook, using CorText

6 CONCLUSION: ONE APPROACH TO GUIDE THEM ALL?

As Marres and Gerlitz (2015) observe, social media platforms 'do not present us with raw data, but rather with specially formatted information' (p.22). The formatting of this data, both at an interface and software (API) level, then inevitably influences the methodological implications for research. By "standardising" the presentation of data and the way it is made visible, the platforms are guiding the researchers through what is available to be seen and investigated. The perspective, methods and insights are limited by the affordances of each and every platform, their algorithmic and API system(s). Keeping this in mind is important for discussing the scope and depth of available information when employing the set of research methods and tools in this empirical research. Marres and Gerlitz (2015) call this 'methodological bias' (p. 22) and rightfully ask the question if "is it really the researcher that here 'decides' to use this method, or is this decision rather informed by the object of study with its associated tools and metrics?" (ibid.). If not limited in the right sense of the word, then we, as researchers, are nudged, steered towards the particular configuration of analytic practices via the platforms', APIs' and software's own 'sampling techniques, options for analysis and modes of visualization' (p. 31).

Another potentially problematic aspect of relying both on APIs and data and being denied access to them and adopting a method for data collection based on observation, is the constant change of what platforms make available. This highlights the constant revision and change of their *politics of visibility* and *politics of knowledge* implemented via the changes at an interface and software level. Barrett and Kreiss (2019) call this *platform transience* – a concept they use to describe the sudden changes platforms make in their policies, procedures and/or affordances, which impacts the ability for critical research, as it makes them continuously changeable and ephemeral in significant ways. Right after the end of the data collection phase of this research, *Facebook* changed its data and ad explanation structures, now offering more information at the disposal of users (Facebook Newsroom, 2019). This is not just problematic in the sense that it makes data collected at different time-periods potentially incomparable, but it also makes the study of algorithmic systems almost a study of 'historical objects' (Bucher, 2012). We as researchers will be always bounded by what platforms decide to make available, either via the interface or the API. With platforms closing their APIs and giving data access only to the "chosen" few (for example, *Facebook's Social Science One*[27]), move described by Bruns (2019) as 'corporate data philanthropy', the data access gap will

[27] More about the *Facebook*-Academic partnership can be read at the following link: https://socialscience.one/. Last accessed July 31, 2020.

be only widening. Hence our ability to study technology, society, and the intersection of the two, will narrow down and become potentially very limited.

Considering this, and considering the increasing limitations of how research can be done and what can be obtained as valuable knowledge, as a result of immanent methodological bias, API restrictions and impenetrability of black boxes, we are faced with the question of how successful and valid the research we conducted was. At an interface level, the methodological design imposed some limitations in a particular manner. This is once again related with how the platforms organise their information: how *Facebook* and *Google's* ad settings are organised, how much they reveal, how the data obtained via Data Subject Access Request is organised, how readable it is and finally, what is made available through these "interfaces" and what is concealed, left out or not provided. Another related aspect concerns the nature of observation as a research method and of the auto-technographic approach. As outlined by Weltevrede (2016), this is always a *me-centric* view, highly individualised and personalised (p. 107), as is our experience and content provided on these platforms. Additionally, we need to be aware of the complexities arising when one would like to translate this very same methodological design and setting on a sample comprising more than one research subject. That would require undertaking additional and modified steps, setting up the research environment and testing the possibilities to obtain valuable and valid data, considering all the complexities of browsing histories, browsing habits and patterns, that particular research subjects could exhibit.

These exact same limitations can be seen as an advantage, as they enable 'real user-algorithmic agent interactions' (Bodo et al., 2018, p. 143). Being able to observe these enriches the quality of the insights, but more importantly, it allows to see the wider 'socio-technological assemblage' (ibid.) and the networks between different actors. And while it might not provide a picture of the totality of the system, it does provide a valuable, although partial, reconstruction of the complexity of these algorithmic assemblages.

By using the affordances of the different methods, at a different level of visibility (interface and software) for analytical inquiry, and combining these findings, new and more in-depth insights were made possible. This is reinforced with the action of repurposing objects of/for study — such as the data explanations, ad explanations, data subjects access request and similar — as a strategy to overcome the limitations, uncover and make visible what was previously not *revealable*. While having to adjust to the affordances and thus limitations of methods and tools, this research and methodological strategy offered ways to be innovative, to — by learning what is possible — look for new avenues, new perspectives, new sources of data and thus insights for digital social research. In that regard, the methodological design of this research is successful, as it provides access to new insights and enables for a more in-depth inquiry into the processes of algorithmic construction of identity, data extraction and inferential analytics, and the ecosystem of actors and networks around these surveillance practices. At a software level, automated tools enabled for a more in-depth knowledge and helped better investigate aspects hidden from the interface and the eye. However, the approach has its limitations, emanating from the nature of platforms' APIs, which are also limited in scope and applicability by their very affordances. They have their own "politics of visibility", limiting what can be seen and uncovered. At an interface level, the daily, detailed observation and recording of the workings and outputs of the system enable for more granular insights and observations of the subtle changes in and by algorithmic systems.

With our research we tried to manoeuvre around the restrictions for research imposed by APIs and black boxes and find ways to investigate opaque algorithmic systems. Following Paßmann and Boersma's (2017) suggestion for pursuing practical transparency, complemented by what they call formalized transparency, we made use of sources external to the algorithms, their APIs and black boxes as a way to detect and make known the unknowns. While APIs are important research entry point, they are not the only one. We experimented with different approaches to circumvent the limitations for research imposed by platforms' gatekeeping practices. In doing so, we got close to what can be called 'digital fieldwork' (Venturini and Rogers, 2019): exploring, experimenting with, testing and employing various new approaches, sources, ways to collect data and capture the interactions between the algorithms and users, mediated via interfaces and APIs. With that, we proposed (just) one of the possible avenues for overcoming data access gaps and algorithmic opacity in doing digital social research. While the question of *if* and *how* platforms should provide access to data for researchers is not a focus of this paper, it remains an important one. We are on the opinion that while it is necessary, thorough digital social research should use and rely on other methods, techniques and data access points in combination with API data. We see this as the only approach that will provide comprehensive view of the socio-technological assemblages, their outputs and impact.

FUNDING STATEMENT AND ACKNOWLEDGMENTS

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was in part supported by Fonds voor Wetenschappelijk Onderzoek – Vlaanderen (FWO) [grant number G054919N]. Ana Pop Stefanija was an Erasmus

Mundus Joint Master Degree scholarship holder, financed by the European Union, while working on part of this research.

REFERENCES

AdAnalyst: Bringing transparency to Facebook Ads. (2019, June 12). https://adanalyst.mpi-sws.org/#about-transpad

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Andrejevic, M., & Gates, K. (2014). Big Data Surveillance: Introduction. *Surveillance & Society*, *12*(2), 185–196. https://doi.org/10.24908/ss.v12i2.5242

Andreou, A., Venkatadri, G., Goga, O., Gummadi, K. P., Loiseau, P., & Mislove, A. (2018). *Investigating ad transparency mechanisms in social media: A case study of Facebook's explanations*. http://www.eurecom.fr/publication/5414

Angwin, J., Mattu, S., Paris, Jr. Terry. (2016, December 27). *Facebook Doesn't Tell Users Everything It Really Knows About Them* [Text/html]. ProPublica. https://www.propublica.org/article/facebook-doesnt-tell-users-everything-it-really-knows-about-them

Ariana Tobin, J. B. M. (2018, September 18). *Facebook Is Letting Job Advertisers Target Only Men* [Text/html]. ProPublica. https://www.propublica.org/article/facebook-is-letting-job-advertisers-target-only-men

Barrett, B., & Kreiss, D. (2019). Platform transience: Changes in Facebook's policies, procedures, and affordances in global electoral politics. *Internet Policy Review*, *8*(4). https://policyreview.info/articles/analysis/platform-transience-changes-facebooks-policies-procedures-and-affordances-global

Bashyakarla, V. (2018a, May 18). *Psychometric Profiling: Persuasion by Personality in Elections*. Our Data Our Selves. https://ourdataourselves.tacticaltech.org/posts/psychometric-profiling/

Bashyakarla, V. (2018b, September 11). *Geotargeting: The Political Value of Your Location*. Our Data Our Selves. https://ourdataourselves.tacticaltech.org/posts/geotargeting/

BBC News. (2018, October 23). *Mobile app data sharing "out of control."* https://www.bbc.com/news/technology-45952466

Beckett, L. (2014, June 13). *Everything We Know About What Data Brokers Know About You* [Text/html]. ProPublica.

https://www.propublica.org/article/everything-we-know-about-what-data-brokers-know-about-you

Better. (2019, July 18). *Trackers Collection*. https://better.fyi/trackers/

Binns, R., Lyngs, U., Van Kleek, M., Zhao, J., Libert, T., & Shadbolt, N. (2018). Third Party Tracking in the Mobile Ecosystem. *Proceedings of the 10th ACM Conference on Web Science*, 23–31. https://doi.org/10.1145/3201064.3201089

Binns, R., Zhao, J., Kleek, M. V., & Shadbolt, N. (2018). Measuring Third-party Tracker Power Across Web and Mobile. *ACM Trans. Internet Technol.*, *18*(4), 52:1–52:22. https://doi.org/10.1145/3176246

Bodo, B., Helberger, N., Irion, K., Borgesius, K. Z., Moller, J., Velde, B. van de, Bol, N., Es, B. van, & Vreese, C. de. (2018). Tackling the Algorithmic Control Crisis -the Technical, Legal, and Ethical Challenges of Research into Algorithmic Agents. *Yale Journal of Law and Technology*, *19*(1), 133–180.

Boerman, S. C., Kruikemeier, S., & Borgesius, F. J. Z. (2017). Online Behavioral Advertising: A Literature Review and Research Agenda. *Journal of Advertising*, *46*(3), 363–376. https://doi.org/10.1080/00913367.2017.1339368

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*, 662–679.

Bozdag, E. (2013). Bias in Algorithmic Filtering and Personalization. *Ethics and Inf. Technol.*, *15*(3), 209–227. https://doi.org/10.1007/s10676-013-9321-6

Brave. (n.d.). *Marked up copy of Google's RTB "Publisher Verticals."* https://brave.com/wp-content/uploads/2019/01/Google-publisher-verticals-marked-up.pdf

Bruns, A. (2019). After the 'APIcalypse': social media platforms and their fight against critical scholarly research, Information, Communication & Society, 22:11, 1544-1566, DOI: 10.1080/1369118X.2019.1637447

Brusseau, J. (2019). Ethics of identity in the time of big data. *First Monday*, *24*(5). https://doi.org/10.5210/fm.v24i5.9624

Buchanan, R. (2001). Design research and the new learning. Design Issues. 17(4), 3-23.

Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media & Society*, *14*(7), 1164–1180. https://doi.org/10.1177/1461444812440159

Bucher, T. (2016). Neither Black Nor Box: Ways of Knowing Algorithms. In S. Kubitschko & A. Kaun (Eds.), *Innovative Methods in Media and Communication Research* (pp. 81–98). Springer International Publishing. https://doi.org/10.1007/978-3-319-40700-5_5

Bucher, T. (2017). The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society, 20*(1), 30–44. https://doi.org/10.1080/1369118X.2016.1154086

Bucher, T. (2018). *If...Then: Algorithmic Power and Politics*. Oxford University Press.

Bucher, T. (n.d.). *Programmed sociality: A software studies perspective on social networking sites*. 221.

Cabañas, J. G., Cuevas, Á., & Cuevas, R. (2018). Facebook Use of Sensitive Data for Advertising in Europe. *ArXiv:1802.05030 [Cs]*. http://arxiv.org/abs/1802.05030

Cheney-Lippold, J. (2011). A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control. *Theory, Culture & Society*, *28*(6), 164–181. https://doi.org/10.1177/0263276411424420

Crain, M. (2018). The limits of transparency: Data brokers and commodification. *New Media & Society*, *20*(1), 88–104. https://doi.org/10.1177/1461444816657096

Crawford, K., Lingel, J., & Karppi, T. (2015). Our metrics, ourselves: A hundred years of self-tracking from the weight scale to the wrist wearable device. *European Journal of Cultural Studies*, *18*(4–5), 479–496. https://doi.org/10.1177/1367549415584857

Crawford, K. (2013, April 1). The Hidden Biases in Big Data. *Harvard Business Review*. https://hbr.org/2013/04/the-hidden-biases-in-big-data

Dance, G. J. X., LaForgia, M., & Confessore, N. (2018, December 18). As Facebook Raised a Privacy Wall, It Carved an Opening for Tech Giants. *The New York Times*. https://www.nytimes.com/2018/12/18/technology/facebook-privacy.html

de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the Crowd: The privacy bounds of human mobility. *Nature Research Journal*, *3*, 1376. https://doi.org/10.1038/srep01376

Desjardins, J. (2018, November 19). *Here's What the Big Tech Companies Know About You*. Visual Capitalist. https://www.visualcapitalist.com/heres-what-the-big-tech-companies-know-about-you/

Dias, T., & Natusch, I. (2017, November 29). They are stalking you to calculate your score. *Chupadados*. https://chupadados.codingrights.org/en/they-are-stalking-you-to-calculate-your-score/

Dieter, M., & Tkacz, N. (2020). The Patterning of Finance/Security: A Designerly Walkthrough of Challenger Banking Apps. *Computational Culture*, *7*. http://computationalculture.net/the-patterning-of-finance-security/

Digital Methods Initiative. (2019, July 27). *The research browser < Dmi < Foswiki*. https://wiki.digitalmethods.net/Dmi/FirefoxToolBar#The_research_browser

Dourish, P. (2016). Algorithms and their others: Algorithmic culture in context. *Big Data & Society*, *3*(2), 2053951716665128. https://doi.org/10.1177/2053951716665128

Facebook. (2019, May 21). *How does Facebook detailed targeting work?* Facebook Ads Help Center. https://www.facebook.com/business/help/182371508761821

Facebook Business. (2019, June 13). *About Lookalike Audiences for Facebook ads*. Facebook Ads Help Center. https://www.facebook.com/business/help/164749007013531

Facebook Business. (2019, June 24). *Lookalike Audiences*. Facebook Business. https://en-gb.facebook.com/business/learn/facebook-ads-lookalike-audiences

Facebook Newsroom. (2019, July 11). *Understand Why You're Seeing Certain Ads and How You Can Adjust Your Ad Experience*. https://newsroom.fb.com/news/2019/07/understand-why-youre-seeing-ads/

Facebook Newsroom. (2020, March 30). *Updating Our Data Access Tools*. https://about.fb.com/news/2020/03/data-access-tools/

Fix AdTech. (2019, February 20). *New evidence filed in RTB complaint*. Fix AdTech. https://fixad.tech/february2019/

Fritsch, K. (2018). Towards an Emancipatory Understanding of Widespread Datafication. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3122269

Ghostery team. (2017). *Tracking the Trackers: Analysing the global tracking landscape with GhostRank*. https://www.ghostery.com/study/

Gillespie, T. (2012, February 2). *Can an Algorithm be Wrong?* Limn. https://limn.it/articles/can-an-algorithm-be-wrong/

Goga, O. (2019, May 15). *Facebook's "transparency" efforts hide key reasons for showing ads*. The Conversation. http://theconversation.com/facebooks-transparency-efforts-hide-key-reasons-for-showing-ads-115790

Golebiewski, M., & Boyd, D. (2018, May 11). Data Voids: Where Missing Data Can Easily Be Exploited. *Data & Society*. https://datasociety.net/output/data-voids-where-missing-data-can-easily-be-exploited/

Google Blog (2018, June 14) *Greater transparency and control over your Google ad experience*. https://blog.google/technology/ads/greater-transparency-and-control-over-your-google-ad-experience/

Gutwirth, S., & De Hert, P. (2008). Regulating Profiling in a Democratic Constitutional State. In M. Hildebrandt & S. Gutwirth (Eds.), *Profiling the European Citizen: Cross-Disciplinary Perspectives* (pp. 271–302). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6914-7_14

Haggerty, K. D., & Ericson, R. V. (2000). The surveillant assemblage. *The British Journal of Sociology*, *51*(4), 605–622. https://doi.org/10.1080/00071310020015280

Hill, K. (2018, September 26). *Facebook Is Giving Advertisers Access to Your Shadow Contact Information*. Gizmodo. https://gizmodo.com/facebook-is-giving-advertisers-access-to-your-shadow-co-1828476051

Jackson, V., Rosenberg, D., Williams, T. D., Brine, K. R., Poovey, M., Stanley, M., Garvey, E. G., PhD, M. K., Raley, R., Ribes, D., Jackson, S. J., & Bowker, G. C. (2013). *"Raw Data" Is an Oxymoron* (L. Gitelman, Ed.). The MIT Press.

Jarrett, K. (2014). A Database of Intention? In R. König & M. Rasch (Eds.), *Society of the query reader: Reflections on web search*. Inst. of Network Cultures.

Just, N., & Latzer, M. (2017). Governance by algorithms: Reality construction by algorithmic selection on the Internet. *Media, Culture & Society*, *39*(2), 238–258. https://doi.org/10.1177/0163443716643157

Kaltheuner, F. (2018, November 7). *I asked an online tracking company for all of my data and here's what I found*. Privacy International. http://privacyinternational.org/feature/2433/i-asked-online-tracking-company-all-my-data-and-heres-what-i-found

Kien, G. (2008). Technography = Technology + Ethnography: An Introduction. *Qualitative Inquiry*, *14*(7), 1101–1109. https://doi.org/10.1177/1077800408318433

Koebler, J., & Maréchal, D. N. (2018, November 16). Targeted Advertising Is Ruining the Internet and Breaking the World. *Motherboard*. https://motherboard.vice.com/en_us/article/xwjden/targeted-advertising-is-ruining-the-internet-and-breaking-the-world

Lapowsky, I., & Thompson, N. (2019, March 6). Facebook's Pivot to Privacy Is Missing Something Crucial. *Wired*. https://www.wired.com/story/facebook-zuckerberg-privacy-pivot/

Lehtiniemi, T. (2019). *Reorienting Datafication? New Roles For Users In Online Platform Markets | The Internet, Policy & Politics Conferences*. 2016. http://blogs.oii.ox.ac.uk/ipp-conference/2016/programme-2016/track-c-markets-and-labour/digital-markets-and-currencies/tuukka-lehtiniemi-reorienting.html

Lev-Aretz, Y. (2019, April). Facebook and the perils of a personalized choice architecture. *TechCrunch*.

http://social.techcrunch.com/2018/04/24/facebook-and-the-perils-of-a-personalized-choice-architecture/

Lindgren, S. (2019). Hacking Social Science for the Age of datafication. *Journal of Digital Social Research, 1(1), 1-9*.

Lyon, D. (2018). *The Culture of Surveillance: Watching as a Way of Life* (1 edition). Polity.

Maass, M., Wichmann, P., Pridöhl, H., & Herrmann, D. (2017). PrivacyScore: Improving Privacy and Security via Crowd-Sourced Benchmarks of Websites. *ArXiv:1705.05139 [Cs]*, *10518*, 178–191. https://doi.org/10.1007/978-3-319-67280-9_10

Mahieu, R. L. P., Asghari, H., & Eeten, M. van. (2018). Collectively exercising the right of access: Individual effort, societal effect. *Internet Policy Review*, *7*(3). https://policyreview.info/articles/analysis/collectively-exercising-right-access-individual-effort-societal-effect

Mai, J.-E. (2016). Big data privacy: The datafication of personal information. *The Information Society*, *32*(3), 192–199. https://doi.org/10.1080/01972243.2016.1153010

Maiberg, E., & Grauer, Y. (2018, March 27). What Are "Data Brokers," and Why Are They Scooping Up Information About You? *Vice*. https://www.vice.com/en_us/article/bjpx3w/what-are-data-brokers-and-how-to-stop-my-private-data-collection

Maiberg, E., Koebler, J., & Cox, J. (2018, December 5). Internal Documents Show Facebook Has Never Deserved Our Trust or Our Data. *Motherboard*. https://motherboard.vice.com/en_us/article/7xyenz/internal-documents-show-facebook-has-never-deserved-our-trust-or-our-data

Mann, M., & Matzner, T. (2019). Challenging algorithmic profiling: The limits of data protection and anti-discrimination in responding to emergent discrimination. *Big Data & Society*, *6*(2), 2053951719895805. https://doi.org/10.1177/2053951719895805

Marres, N., & Gerlitz, C. (2016). Interface Methods: Renegotiating relations between digital social research, STS and sociology. *Sociological Review*, *64*, 21–46.

Marres, N. (2017). Digital Sociology. Cambridge: Polity Press.

Masson, E., van Es, K., Wieringa, M., (2020). Data Walking for Critical Data Studies: An Explorative Survey of Walking Methodologies. *Digital Culture & Education*, 11(1), 36-52

Matzner, T. (2016). Beyond data as representation: The performativity of Big Data in surveillance. *Surveillance & Society*, *14*(2), 197–210. https://doi.org/10.24908/ss.v14i2.5831

Milan, S. (2018). Digital Traces in Context| Political Agency, Digital Traces, and Bottom-Up Data Practices. *International Journal of Communication*, *12*(0), 21.

Milan, Stefania, & van der Velden, L. (2016). The Alternative Epistemologies of Data Activism. *Digital Culture & Society*, *2*(2), 57–74. http://dx.doi.org/10.25969/mediarep/991

Milan, Stefanija, & Agosti, C. (2019, February 7). *Personalisation algorithms and elections: Breaking free of the filter bubble*. Internet Policy Review. https://policyreview.info/articles/news/personalisation-algorithms-and-elections-breaking-free-filter-bubble/1385

Mittelstadt, B. (2016). Auditing for Transparency in Content Personalization Systems. *International Journal of Communication*, *10*(0), 12.

Munk, A. K. (2019, February 20). Introduction to semantic analysis with Cortext. Retrieved June 21, 2019, from Anders Kristian Munk website: https://medium.com/@AnthropologicalMachines/introduction-to-semantic-analysis-with-cortext-19f355b7289a

Neyland, D. (2016). Bearing Account-able Witness to the Ethical Algorithmic System. *Science, Technology, & Human Values*, *41*(1), 50–76. https://doi.org/10.1177/0162243915598056

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

Olejnik, L., Castelluccia, C., & Janc, A. (2014). On the uniqueness of Web browsing history patterns. *Annals of Telecommunications - Annales Des Télécommunications*, *69*(1), 63–74. https://doi.org/10.1007/s12243-013-0392-5

Papadopoulos, P., Kourtellis, N., & Markatos, E. P. (2018). *Cookie Synchronization: Everything You Always Wanted to Know But Were Afraid to Ask*. https://doi.org/10.1145/3308558.3313542

Pariser, E. (2012). *The Filter Bubble: What The Internet Is Hiding From You*. Penguin.

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.

Paßmann, J., & Boersma, A. (2017). Unknowing Algorithms. In K. van Es & M. T. Schäfer (Eds.), *The Datafied Society. Studying Culture through Data* (pp. 139–146). Amsterdam University Press. https://doi.org/10.5117/9789462981362

Pasternack, S. M. and A. (2019, March 2). *Here are the data brokers quietly buying and selling your personal information*. Fast Company. https://www.fastcompany.com/90310803/here-are-the-data-brokers-quietly-buying-and-selling-your-personal-information

Pierson, J. (2014). Interdisciplinary Perspective on Social Media, Privacy and Empowerment: The Role of Media and Communication Studies in Technological Privacy Research. *Digital Enlightenment Yearbook 2014*, 265–274.

Privacy International. (2019, April 16). *How Apps on Android Share Data with Facebook—Report*. Privacy International. http://privacyinternational.org/report/2647/how-apps-android-share-data-facebook-report

Raley, R. (2013). Dataveillance and Surveillance. In *"Raw Data" is an Oxymoron* (p. 192). The MIT Press. https://mitpress.mit.edu/books/raw-data-oxymoron

Reigeluth, T. B. (2014). Why data is not enough: Digital traces as control of self and self-control. *Surveillance & Society*, *12*(2), 243–254. https://doi.org/10.24908/ss.v12i2.4741

Rieder, B., & Röhle, T. (2017). Digital Methods. From Challenges to Bildung. In K. van Es & M. T. Schäfer (Eds.), *The Datafied Society. Studying Culture through Data* (pp. 109–124). Amsterdam University Press. https://doi.org/10.5117/9789462981362

Rocher, L., Hendrickx, J. M., & Montjoye, Y.-A. de. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications*, *10*(1), 3069. https://doi.org/10.1038/s41467-019-10933-3

Roesner, F., Kohno, T., & Wetherall, D. (2012). *Detecting and Defending Against Third-Party Tracking on the Web*. 155–168. https://www.usenix.org/conference/nsdi12/technical-sessions/presentation/roesner

Roesner, F., Rovillos, C., Saxena, A., & Kohno, T. (2019, June 6). *TrackingObserver: A Browser-Based Web Tracking Detection Platform*. https://trackingobserver.cs.washington.edu/

Rogers, R. (2013). *Digital Methods*. MIT Press.

Rogers, R. (2015). Digital Methods for Web Research. In *Emerging Trends in the Social and Behavioral Sciences* (pp. 1–22). American Cancer Society. https://doi.org/10.1002/9781118900772.etrds0076

Rogers, R. (2017). Foundations of digital methods: Query design. In K. van Es & M. T. Schäfer (Eds.), *The Datafied Society. Studying Culture through Data* (pp. 125–137). Amsterdam University Press. https://doi.org/10.5117/9789462981362

Rogers, R. (n.d.). *The Googlization Question, and the Inculpable Engine*.

Rogers, R. (2018). Social Media Research After the Fake News Debacle. *PARTECIPAZIONE E CONFLITTO*, *11*(2), 557-570–570. https://doi.org/10.1285/i20356609v11i2p557

Ryan, J. (2018, September 12). *Regulatory complaint concerning massive, web-wide data breach by Google and other "ad tech" companies under Europe's*

GDPR. Brave Browser. https://www.brave.com/blog/adtech-data-breach-complaint/

Ryan, J. (2018). *Behavioural advertising and personal data*. Brave. https://brave.com/Behavioural-advertising-and-personal-data.pdf

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

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014, May 22). *Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms*. 64th Annual Meeting of the International Communication Association, Seattle, WA, USA. https://pdfs.semanticscholar.org/b722/7cbd34766655dea10d0437ab10df3a127396.pdf

Schmidt, C. D. (2018). *Google data collection research* (pp. 1–55). Digital Content Next, Vanderbilt University. https://digitalcontentnext.org/wp-content/uploads/2018/08/DCN-Google-Data-Collection-Paper.pdf

Schwartz, O. (2018, July 13). *Digital ads are starting to feel psychic*. The Outline. https://theoutline.com/post/5380/targeted-ad-creepy-surveillance-facebook-instagram-google-listening-not-alone

Seaver, N. (2013). Knowing algorithms. *Media in Transition 8*, 1–12. https://static1.squarespace.com/static/55eb004ee4b0518639d59d9b/t/55ece1bfe4b030b2e8302e1e/1441587647177/seaverMiT8.pdf

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, *4*(2), 2053951717738104. https://doi.org/10.1177/2053951717738104

SOCIAL SCIENCE ONE. (n.d.). Retrieved August 6, 2020, from https://socialscience.one/home

Store, P. D. (2016, December 27). *Facebook Ad Categories* [Text/html]. ProPublica Data Store. https://www.propublica.org/datastore/dataset/facebook-ad-categories

Sweeney, L. (2013). Discrimination in Online Ad Delivery. *COMMUNICATIONS OF THE ACM*, *56*(5), 44–54.

Szymielewicz, K. (2019, January 30). *Legal battle over online behavioural advertising widening*. Internet Policy Review. https://policyreview.info/articles/news/legal-battle-over-online-behavioural-advertising-widening/1384

Taylor, A., & Sadowski, J. (2015, May 27). *How Companies Turn Your Facebook Activity Into a Credit Score*. https://www.thenation.com/article/how-companies-turn-your-facebook-activity-credit-score/

Thompson, S. A. (2019, April 30). Opinion | These Ads Think They Know You. *The New York Times*. https://www.nytimes.com/interactive/2019/04/30/opinion/privacy-targeted-advertising.html, https://www.nytimes.com/interactive/2019/04/30/opinion/privacy-targeted-advertising.html

Tufekci, Z. (2019, March 8). Opinion | Zuckerberg's So-Called Shift Toward Privacy. *The New York Times*. https://www.nytimes.com/2019/03/07/opinion/zuckerberg-privacy-facebook.html

Uricchio, W. (2017). Data, Culture and the Ambivalence of Algorithms. In K. van Es & M. T. Schäfer (Eds.), *The Datafied Society. Studying Culture through Data* (pp. 125–137). Amsterdam University Press. https://doi.org/10.5117/9789462981362

van Dijck, José. (2013). 'You have one identity': Performing the self on Facebook and LinkedIn. *Media, Culture & Society*, *35*(2), 199–215. https://doi.org/10.1177/0163443712468605

van Dijck, Jose. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society, 12*(2), 197–208. https://doi.org/10.24908/ss.v12i2.4776

van Dijck, J., & Poell, T. (2013). Understanding Social Media Logic. *Media and Communication*, *1*(1), 2–14. https://doi.org/10.17645/mac.v1i1.70

van Es, K., & Schäfer, M. T. (2017). Introduction. New Brave World. In K. van Es & M. T. Schäfer (Eds.), *The Datafied Society. Studying Culture through Data* (pp. 15–22). Amsterdam University Press. https://doi.org/10.5117/9789462981362

van Es, K., López Coombs, N., & Boeschoten, T. (2017). Towards a Reflexive Digital Data Analysis. In K. van Es & M. T. Schäfer (Eds.), *The Datafied Society. Studying Culture through Data* (pp. 171–180). Amsterdam University Press. https://doi.org/10.5117/9789462981362

van Es, K., Wieringa, M., and Schafer, M.T. (2018). Tool Criticism: From Digital Methods to Digital Methodology. In International Conference on Web Studies (WS.2 2018), October 3-5, 2018, Paris, France. ACM, New York, NY, USA.

Venkatadri, G., Andreou, A., Liu, Y., Mislove, A., Gummadi, K. P., Loiseau, P., & Goga, O. (2018). Privacy Risks with Facebook's PII-Based Targeting: Auditing a Data Broker's Advertising Interface. *2018 IEEE Symposium on Security and Privacy (SP)*, 89–107. https://doi.org/10.1109/SP.2018.00014

Venkatadri, G., Sapiezynski, P., Redmiles, E., Mislove, A., Goga, O., Mazurek, M., & P. Gummadi, K. (2019). Auditing Offline Data Brokers via Facebook's Advertising Platform. *Proceedings of the 2019*

World Wide Web Conference (WWW'19), 1920–1930. https://doi.org/10.1145/3308558.3313666

Venkatadri, Giridhari, Lucherini, E., Sapiezynski, P., & Mislove, A. (2019). Investigating sources of PII used in Facebook's targeted advertising. *Proceedings on Privacy Enhancing Technologies*, *2019*(1), 227–244. https://doi.org/10.2478/popets-2019-0013

Venturini, T., Meunier, A (2019). Drafting and atlas on Artificial Intelligence's matters of reflection. Available at: http://www.tommasoventurini.it/ai/ . Last accessed January 29, 2020.

Venturini, T., Rogers, R. (2019). "'API-Based Research' or How Can Digital Sociology and Digital Journalism Studies Learn from the Cambridge Analytica Affair." Digital Journalism, Forthcoming.

Venturini, T., Bounegru, L., Gray, J., & Rogers, R. (2018). A reality check(list) for digital methods. *New Media & Society*, *20*(11), 4195–4217. https://doi.org/10.1177/1461444818769236

Wachter, S. (2019). *Affinity Profiling and Discrimination by Association in Online Behavioural Advertising* [SSRN Scholarly Paper]. Social Science Research Network. https://papers.ssrn.com/abstract=3388639

Wachter, S., & Mittelstadt, B. (2018). *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI* (SSRN Scholarly Paper No. ID 3248829). Retrieved from Social Science Research Network website: https://papers.ssrn.com/ abstract=3248829

Weltevrede, E. J. T. (2016). *Repurposing digital methods: The research affordances of platforms and engines* [University of Amsterdam]. https://dare.uva.nl/search?identifier=aaaa9bb3-8647-41df-954c-2bb1e9f15d77

WhoTracksMe. (2019, July 18). *Trackers Rank*. https://whotracks.me/trackers.html

Xu, Y. (Joe). (2018). Programmatic Dreams: Technographic Inquiry into Censorship of Chinese Chatbots. *Social Media + Society*, *4*(4), 2056305118808780. https://doi.org/10.1177/2056305118808780

Zimmer, M. (2008). *The Gaze of the Perfect Search Engine: Google as an Infrastructure of Dataveillance* (A. Spink & M. Zimmer, Eds.; pp. 77–99). https://doi.org/10.1007/978-3-540-75829-7_6

Zuboff, S. (2015). Big other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology*, *30*(1), 75–89. https://doi.org/10.1057/jit.2015.5

Zuboff, S. (2016, March 5). Google as a Fortune Teller: The Secrets of Surveillance Capitalism. *Frankfurter Allgemeine Zeitung*. https://www.faz.net/1.4103616

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile Books.

JOURNAL OF DIGITAL SOCIAL RESEARCH

WWW.JDSR.IO ISSN: 2003-1998

VOL. 2, NO. 3, 2020, 126–157

ARTIFICIAL INTELLIGENCE AND VIDEO GAME CREATION: A FRAMEWORK FOR THE NEW LOGIC OF AUTONOMOUS DESIGN

Stefan Seidel[a], Nicholas Berente[b], Aron Lindberg[c], Kalle Lyytinen[d], Benoit Martinez[e], and Jeffrey V. Nickerson[c]

ABSTRACT

Autonomous, intelligent tools are reshaping all sorts of work practices, including innovative design work. These tools generate outcomes with little or no user intervention and produce designs of unprecedented complexity and originality, ushering profound changes to how organizations will design and innovate in future. In this paper, we formulate conceptual foundations to analyze the impact of autonomous design tools on design work. We proceed in two steps. First, we conceptualize autonomous design tools as 'rational' agents which will participate in the design process. We show that such agency can be realized through two separate approaches of information processing: symbolic and connectionist. Second, we adopt control theory to unpack the relationships between the autonomous design tools, human actors involved in the design, and the environment in which the tools operate. The proposed conceptual framework lays a foundation for studying the new kind of material agency of autonomous design tools in organizational contexts. We illustrate the analytical value of the proposed framework by drawing on two examples from the development of Ubisoft's *Ghost Recon Wildlands* video game, which relied on such tools. We conclude this essay by constructing a tentative research agenda for the research into autonomous design tools and design work.

Keywords: autonomous design tools, artificial intelligence, organizing, design, innovation, digital innovation, control, work

[a] University of Liechtenstein, Liechtenstein
[b] University of Notre Dame, Indiana, USA
[c] Stevens Institute of Technology, New Jersey, USA
[d] Case Western Reserve University, Ohio, USA
[e] Ubisoft Paris, France

1 INTRODUCTION

Digital technologies increasingly shape the environments in which they operate (Baskerville, Myers, & Yoo, 2019; Rai, Constantinides, & Sarker, 2019) by acting as "performative material devices" (Pickering, 1995). Performativity implies that digital technologies operate with some level of autonomy. Advanced forms of such technologies possess some form of artificial intelligence (AI). Such technologies have information processing capabilities for transforming some inputs into outputs in a way that can be deemed intelligent without close human monitoring. As a result, they have the genuine "capacity … to act on their own, apart from human intervention" (Leonardi, 2011, p. 148). Such autonomy is now evident in a growing array of technologies, including self-driving cars (Badue et al., 2019), conversational agents or chatbots (Cassell, Sullivan, Churchill, & Prevost, 2000), Internet of Things (IoT) applications such as smart homes (Porter & Heppelmann, 2014), and indeed autonomous design tools (Shaker, Togelius, & Nelson, 2016). These technologies can, to an extent, act on their own with little or no human intervention and in ways that are not fully predictable or understandable by humans. These tools also shape their environment in multifaceted ways. Therefore, it is no longer appropriate to view these technologies as passive inert entities to be enacted by humans as controllable tools.

This development has been fueled, in part, by the increased use of AI techniques, such as machine learning or genetic algorithms. These techniques have been evolving for decades in the AI community but have only recently become more widespread and productive in organizational settings (Daugherty & Wilson, 2018). The increased deployment of such autonomous tools has been fueled by effective access to large swaths of data and computing power enabled by the emergence of broadband networks, sensor technologies, cloud-based computing, and platform induced ecosystems (Parker, Van Alstyne, & Jiang, 2016; Tiwana, 2015). As a result, many digital applications are no longer merely passive tools that support or control manual tasks and related organizational processes. They are no longer systems that merely automate a pre-defined process and then 'informate' that process (Zuboff, 1988; Seidel & Berente 2020). In addition, many systems can now act in ways that were previously reserved for human agents (Lyytinen, Nickerson, & King, 2020; Seidel, Berente, Lindberg, Nickerson, & Lyytinen, 2019). This shift has given rise to new concepts, typologies, and notions, such as machines as teammates (Seeber et al., 2019), human-machine-learning (Seidel, Berente, Lindberg, et al., 2019), role-reversal (Demetis & Lee, 2017), digital agency (Ågerfalk, 2020), and meta-human systems (Lyytinen et al., 2020). Humans now delegate tasks to tools that act with autonomy (Ågerfalk, 2020; Zhang, Yoo, Lyytinen,

& Lindberg, forthcoming). Autonomously acting intelligent, learning algorithms increasingly make decisions and engage in value judgements (Baskerville et al., 2019). They rely on their own percepts instead of just executing upon prior knowledge conveyed by their designers (Russel & Norvig, 2016). In some situations, autonomous systems can be conceived of as "users" of humans, rather than the reverse (Baskerville et al., 2019; Demetis & Lee, 2017; Lyytinen et al., 2020). These developments call upon a posthumanist lens that does not identify humans as the sole sources of agency, but considers human and material agencies on equal footing and in symbiotic relationships with a circumscribed sociotechnical system (Latour, 2005; Pickering, 1993, 1995).

One domain that has openly embraced software with autonomous capabilities and epitomizes processes that have traditionally been viewed as human-centric is that of design. Designers across industries increasingly use software-based systems that make independent design decisions. In some cases, these systems execute entire design processes to generate artifacts of ever greater complexity (Seidel, Berente, Lindberg, et al., 2019). Such autonomous design tools employ multiple computational approaches to generate design artifacts, including path-finding algorithms, meta-heuristics (in particular, evolutionary algorithms), and neural networks. Using such techniques, autonomous design tools can now generate a growing variety of multifaceted design artifacts, for instance, nearly full designs of next-generation computer chips (Brown & Linden, 2011; Zhang et al., forthcoming), user interfaces (Yumer, Asente, Mech, & Kara, 2015), three-dimensional virtual worlds (Smelik, Tutenel, de Kraker, & Bidarra, 2010b), and static as well as dynamic content for video games and feature films (Hendrikx, Meijer, Van Der Velden, & Iosup, 2013; Togelius, Yannakakis, Stanley, & Browne, 2011). The applications for such systems are now expanding to mechanical engineering, aerospace, and architecture, among others.

Empirical evidence suggests that autonomous design tools are fundamentally changing the organizing of innovative design work and the way that designers[1] will generate artifacts in the future (Seidel et al., 2018; Zhang et al., forthcoming). Instead of creating artifacts by directly manipulating multifaceted design representations, designers will increasingly focus on selecting system goals, features, and constraints, deciding on related design parameters, setting values for these parameters, and evaluating and learning from the analysis of the tool outcomes (Seidel, Berente, Lindberg, et al., 2019; Seidel et al., 2018; Summerville et al., 2018). Design work in such environments requires designers to be mindful of the

[1] Note that we use the term "designer" in its broadest sense, to refer to engineers, developers, architects, etc., that draw on their expertise to generate solutions.

logic, capabilities, and limitations of the deployed algorithms and to find ways to make sense of and deal with complex and unanticipated outputs. This opens up important questions related to organizing design, including problems of coordination, control and learning in design teams (Puranam, Alexy, & Reitzig, 2014; Seidel, Berente, Lindberg, et al., 2019).

Design automation—autonomous or otherwise—has significantly improved the efficiency of design across a variety of fields. One could easily conceive autonomous design tools simply as the next wave of automation. Indeed, the literature on the algorithms that generate artifacts often highlights the significant potential of these tools to automate design and introduce scale efficiencies (e.g., Smelik, Tutenel, de Kraker, & Bidarra, 2010a; Togelius et al., 2011). This is a reasonable position—designers use the tools first to automate parts of current design practices by carrying out algorithmically specific, relatively complex pre-programmed tasks (such as wiring between gates in chip design). However, these tools will increasingly also make design decisions that are, at least partially, independent of and not fully knowable to the designer. In other words, the tools become black boxed and start acting autonomously. They carry out many tasks with unprecedented speed, scale, and scope so that these activities are likely to materially change the way designers generate artifacts (Summerville et al., 2018). They also exhibit capacities that fundamentally differ from past computer aided design (CAD) tools supporting manual design activities of architects and engineers (Chang & Wysk, 1997; Gupta, Garg, & Chadha, 1981).

Against this backdrop we posit that the use of autonomous tools will continue to generate profound changes in how organizations design, innovate, and organize related activities. **The aim of this paper, therefore, is to formulate a conceptual framework that facilitates future inquiries into how the new and changed material agency of autonomous design tools shapes organizational contexts, how these tools interact with their environment, and how their deployment is likely to lead to novel design processes and artifacts.** To this end, we first conceptualize autonomous design tools a 'rational agents' (Russel & Norvig, 2016) with an embedded design model realized through two separate approaches of information processing: symbolic and connectionist. In a second step, we draw on control theory (Mesarovic, Macko, & Takahara, 1970) to spell out the relationships between autonomous design tools, human designers, and the environment in which the tools are used. At this, we highlight how the need for delegation as well as the frame problem (Dennett, 2006; McCarthy & Hayes, 1981) provide explanations for why control units such as human designers are necessary in typical design situations. We illustrate the analytic value of our model by using two examples adopted from the production of a complex video game software—Ubisoft's *Ghost Recon*

Wildlands. We summarize how autonomous design tools are likely to change the organization of design work in many walks of design given the access to new types of human-machine configurations that are now emerging. We also note avenues for future research on autonomous design tools.

2 AUTONOMOUS DESIGN TOOLS

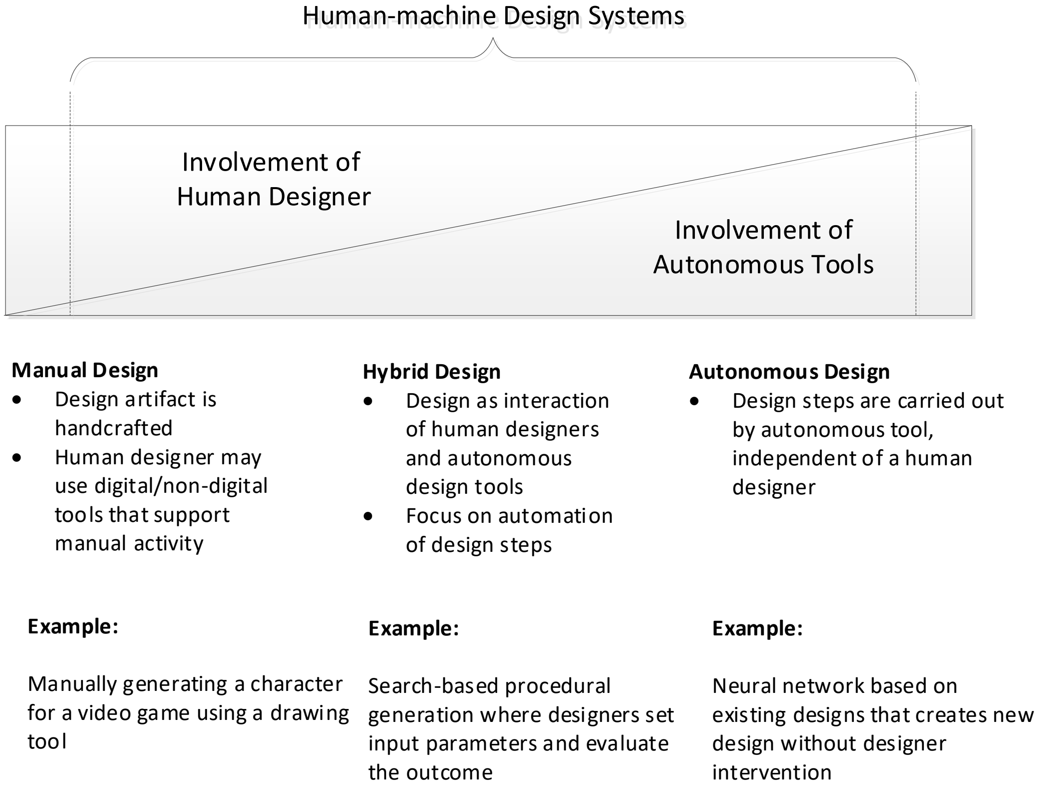
2.1 From Manual Design to Autonomous Design

Design, in the most general sense of the word, involves the formulation of desirable future states in the world (Goel & Pirolli, 1992). To design is to devise "courses of action aimed at changing existing situations into preferred ones" (Simon, 1996, p. 111). Design is simultaneously mental and representational (Baxter & Berente, 2010; Gero, 1990; Goel & Pirolli, 1992). As result, design processes synthesize a solution by iteratively mobilizing and integrating diverse knowledge elements into varied representations of a solution. Design involves exploration and decomposition, as well as synthesis (Goel & Pirolli, 1992). The outcome of design is an artifact—an object generated by human ingenuity and meeting the goal of changing the given situation to a preferred one.

We can broadly classify approaches to design by their utilization of technologies with increasing degrees of autonomy (Figure 1). At one end of the continuum one can find manual design practices where human designers handcraft artifacts. This does not exclude the use of tools that digitize these practices—drawing tools and CAD tools are prominent examples. Here, tools provide detailed affordances for potential actions (Markus & Silver, 2008; Zammuto, Griffith, Majchrzak, Dougherty, & Faraj, 2007) that can be enacted by designers to manipulate and make sense of the representations. Designers are viewed as craftspeople, who, through their deep knowledge of materials, tools, and design principles, intentionally design and shape an artifact (Sennett, 2008). The notion of affordance describes how these tools become involved in design as they express the meaning and intent of the designer to use a tool feature to achieve a specific goal. In design situations affordances are the action potentials that the material properties of a tool offer to some designer or a group of designers (Markus & Silver, 2008). Designers *enact*—that is, they recurrently interact with the technology (Orlikowski, 2000) in their design practices by putting the technology to use; the enacted affordances improve the design performance of the designers who control the tools.

The more technology starts making decisions on behalf of the designer, the more we can conceive the technology as *acting autonomously*.

If one uses technology that makes autonomous decisions, but still involves intermittent interactions with designers, a hybrid human-machine design system is formed. In such a system the degrees of interaction between autonomous systems and human designers will vary significantly. At a minimum, designers state design requirements (goals, constraints) and complete the design by evaluating it against set up performance goals. The focus is still mainly on automating a specific design task such as a placement, composition, or an optimization problem (Summerville et al., 2018; Togelius et al., 2011). At the other end of the continuum we can find fully autonomous tools that create artifacts without a designer's intervention. This is the case where a machine-learned system independently makes all design decisions and can even adjudicate and establish new design goals (Summerville et al., 2018).


Human-machine Design Systems
Involvement of Human Designer
Involvement of Autonomous Tools

Manual Design
- Design artifact is handcrafted
- Human designer may use digital/non-digital tools that support manual activity

Example:

Manually generating a character for a video game using a drawing tool

Hybrid Design
- Design as interaction of human designers and autonomous design tools
- Focus on automation of design steps

Example:

Search-based procedural generation where designers set input parameters and evaluate the outcome

Autonomous Design
- Design steps are carried out by autonomous tool, independent of a human designer

Example:

Neural network based on existing designs that creates new design without designer intervention

Figure 1. The continuum of human-machine design systems (extended from Seidel, Berente, Lindberg, et al., 2019)

While manual design has dominated all areas of design—from arts and architecture, to engineering—we now see an increased deployment of hybrid human-design systems, where design practices involve rich and multifaceted interactions between designers and varied and complex tool sets. Systems used in design having varying degrees of autonomy have been discussed under multiple labels, including procedural generation

(Ashlock & McGuinness, 2013; Hendrikx et al., 2013), procedural modeling (Müller, Wonka, Haegler, Ulmer, & Van Gool, 2006; Parish & Müller, 2001), computational creativity (Liapis, Yannakakis, & Togelius, 2014), generative design systems (Krish, 2011), and autonomous generation (Summerville et al., 2018). What these tools share in common is that they (partially) replace manual craftsmanship in that they generate design artifacts with relatively infrequent designer intervention to find novel solutions that meet given goals and constraints. In this view, autonomous design tools are machine-based agents that perform design work alongside human agents. **Autonomous design tools are software tools that, once started, independently make design decisions to generate design outcomes based on varied forms of input and using an embedded, often complex, unknown, and evolving design model.**

2.2 The Key Elements of Autonomous Design Tools

Autonomous design tools, as defined above, are rational agents. Russel and Norvig (2016) describe rational agents as entities that perceive their environment through sensors, act upon the environment through actuators, and whose behavior can be described in terms of an agent function. In addition, there needs to be some performance measure by which the success of the agent's actions can be evaluated. Rational agents act autonomously to the extent that they rely on their own percepts (the input they receive from the environment) and less on the prior knowledge of their designers (Russel & Norvig, 2016)

We can thus define autonomous design tools by their inputs, their outputs, and, in between, the computational process underlying the specific design decisions the systems make. Embedded design models broadly determine the ways in which the tool will generate outcomes based on a set of input parameters (Seidel, Berente, Lindberg, et al., 2019). That is, as with other information technologies, these tools link inputs to outputs through some form of information processing. The key, however, is that this information processing allows the tool to generate a design outcome by making design decisions that are at least partially independent from human designers (or, more broadly, the users of the tool). From the perspective of the designer interacting with the tool, these outputs can also be unpredictable and surprising. While the designer may have a broad understanding of what the tool is expected to do, he or she cannot precisely anticipate what the tool will produce given the inputs. This is different from mere automation, where a given task is accomplished through a deterministic, traceable process and the designer knows what the output will be given his or her inputs. Consequently, the outcomes generated by autonomous tools are often perceived as being creative by humans (Boden, 2009; Veale, Cardoso, & y Pérez, 2019).

Drawing on Russel and Norvig's (2016) conceptualization of rational agents that have information processing capacity and interact with their environment by receiving sensory input and acting upon that environment, Figure 2 delineates an abstract model of an autonomous design tool. In this view, an autonomous design tool receives sensory input from the environment, makes design decisions based on an embedded design model, and then generates some output that adds content to, or alters, existing design content. Note that the embedded design model can be implemented in various ways, ranging from a simple reflex agent to a learning agent that involves a learning element which allows making improvements based on the model's interaction with the environment (Russel & Norvig, 2016). We next turn to two dominant approaches to implementing embedded design models.

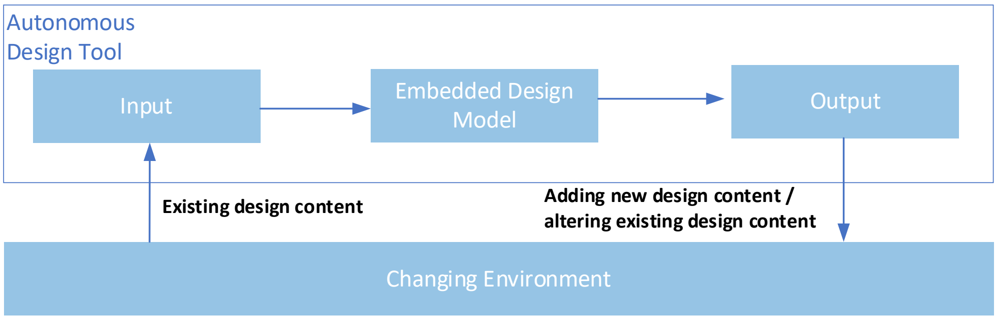


Figure 2. Autonomous tool interacting with its environment (adapted from Russel & Norvig, 2016)

2.3 Two Dominant Types of Embedded Design Models

There are two dominant approaches to the implementation of autonomous design tools—physical symbol systems and non-symbolic connectionist systems. They are based on two main perspectives of information processing (Smolensky, 1987; Sun, 1999). The first is founded on the explicit manipulation of symbol systems expressing the embedded design model based on formal logic. This approach presupposes that the features to be manipulated have already been identified and that consequences of its manipulation can be largely predicted. The second non-symbolic approach is founded on the implicit feature discovery facilitated by connectionist systems, epitomized by artificial neural networks. We can apply these two approaches to distinguish between two broad types of embedded design models (i.e., the computational models that define how the tool works) of autonomous design tools (Table 1). Note that the symbolic vs. non-symbolic categorization is typically maintained in order to distinguish two types of applications in artificial intelligence (e.g. Sun, 1999).

Table 1. Two types of embedded design models

Type of embedded design model	Description	Autonomous design example
Physical symbol system	• Approaches based on explicit representations and symbolic programming • Transformation of physical symbols based on rules • The rules represent a designer's understanding of how the autonomous design tools should address its design approach • Explicit representation of the problem space in terms of relevant features	Search-based algorithms such as pathfinding (Pohl, 1970) in procedural game development (Togelius et al., 2011) Rule-based procedural content generation (Smith & Mateas, 2011)
Non-symbolic/ connectionist systems	• Implicit representation of the problem as the systems discover variables, correlations between variables, and correlations between correlations • Transformation of inputs and outputs through a multilayered network • The underlying representational model is opaque to the designer • No explicit conceptual foundation • Typically based on large data sets ("big data")	Neural network used to reduce complex design problems as in the case of designing user interfaces at Adobe Labs (Yumer et al., 2015) Adversarial networks to generate visual content based on models trained on existent designs (Summerville et al., 2018) Terrain design through adversarial neural networks trained on real-world terrain as well as their sketched counterparts (Guérin et al., 2017)

2.3.1 The Physical Symbol Systems Approach

The physical symbol systems approach is based on the premise that the sort of problem-solving associated with design work is essentially about transforming symbol structures until a result is reached that is satisfactory by some performance measures (Newell & Simon, 1972). In such situations, designers—human or non-human—search a large, multi-dimensional, potentially unbounded, problem space to identify a solution. A problem space is comprised of an initial state, a goal state, and a set of operators that allow a movement from the initial state to the goal state (Newell & Simon, 1972). Designers need a representation of the problem space in order to make it possible to apply operators (Simon, 1996): "A problem representation structures the problem space with elements of the problem and its potential solution and is the most potent explanation for if, and how, a design problem will be solved" (Boland, 2004, p. 106). Throughout the design process, the designer generates design representations that are tested against his or her cognitive schemata for goal satisfaction (Baxter & Berente, 2010). Hence, design can be understood as a search built on nested generate-test cycles that seeks satisfactory solutions for a given, often changing and fluid, design problem (Buchanan, 1992). In this view, optimization is possible, but only in formally constrained and well-structured design situations such as optimal placement of logic gates on a relatively small semiconductor chip, where, for example, optimization techniques such as dynamic programming can be applied. Optimization, however, is a distant or impossible goal in most real-world design situations. The problem spaces are simply too large and complex, and the search takes too much time and effort. While the actual problem space (Dorst & Cross, 2001; Newell & Simon, 1972; Simon, 1996) might be known in an abstract sense (such as in chip design), it is impossible to explore all feasible solutions. Therefore, designers need to employ satisficing procedures and rely on heuristics such as means-ends analysis that involve recursively decomposing the problem into subcomponents until concrete operators can be applied to find solutions that are acceptable rather than optimal (Simon, 1996). Oftentimes design in complex situations adopts procedures that produce initial conditions for further design (Simon, 1996). These procedures can result in a series of component optimization and satisficing actions that ultimately constitute the outcome of the design process.

Typical applications of this approach are based on search-based (Togelius et al., 2011) and rule-based (Smith & Mateas, 2011) approaches for generating design artifacts. Search-based algorithms generate alternative solutions step-by-step and evaluate them. Rule-based approaches apply a set of rules to derive a satisfactory outcome. Tools using these approaches generate large scale artifacts as these tools, upon receiving input from the designers, can normally undertake multiple design steps independently (Ashlock & McGuinness, 2013; Hendrikx et al., 2013).

A prominent application area of these types of design tools is in the procedural generation of content for video games (Ashlock & McGuinness, 2013; Hendrikx et al., 2013). Such content generation can happen at build time (before the game is shipped) and runtime (when the player has started the game). Procedural generation is, in contrast to manual content production, "the application of computers to generate game content, distinguish interesting instances among the ones generated, and select entertaining instances on behalf of the players" (Hendrikx et al., 2013, p. 1:2). Recent well-known examples of the use of build-time procedural content generation can be found in open-world games were users can freely explore a vast virtual environment. These games are based on the availability of large game spaces that would be prohibitively expensive to create without the help of systems that generate large parts of the space without much human designer intervention.

2.3.2 *The Connectionist Approach*

Connectionist approaches, most notably artificial neural networks, provide an alternative. This approach is now used in multiple fields, including design applications (e.g., Yumer et al., 2015). These approaches do not need pre-existing ontologies or features. Instead, such systems discover features from raw sensory data. That is, they do not need pre-existing "theories" and "constructs" to operate; they will discover variables, correlations between variables, and correlations between correlations by themselves. Neural networks can extend the abstraction of such processes layer-by-layer until higher-level constructs in data are discovered, capturing real-world features such as objects, words, and sentences. Because of the mechanisms through which such networks operate, they also are good at compressing information in efficient ways and reducing the dimensionality of large datasets. Through such reduction, design problems can be made more amenable for human designers to navigate a limited set of critical parameters. A key difference from search- or rule-based approaches, which generate content through searching a design space, is that these tools directly generate content (Summerville et al., 2018) in that the systems are trained on successful or representative designs and then can generate other, similar designs (Summerville et al., 2018). In this approach it is not necessary to codify explicit design knowledge in terms of search algorithms that can generate content and then evaluate that content; embedded design models based on connectionist approaches are therefore an important step towards increasing autonomy as they do not rely on the prior knowledge of their designers (Russel & Norvig, 2016).

In the case of designing interfaces at Adobe, for instance, designers were confronted with a problem space that was too large for human designers to navigate—approximately 100 parameters controlled processes for generating navigation structures (Yumer et al., 2015). They turned to creating a deep neural network that helped them to reduce the high-dimensional space to a three-dimensional space that designers could control through slider bars. The designers describe how they used a learning system instead of a rule-based, procedural modeling system to tackle the high dimensionality of the problem as follows:

> Procedural modeling systems allow users to create high quality content through parametric, conditional or stochastic rule sets. While such

> approaches create an abstraction layer by freeing the user from direct geometry editing, the nonlinear nature and the high number of parameters associated with such design spaces result in arduous modeling experiences for non-expert users. We propose a method to enable intuitive exploration of such high dimensional procedural modeling spaces within a lower dimensional space learned through autoencoder network training (Yumer et al., 2015, p. 109).

Symbolic and connectionist approaches for generating design outcomes can also be combined. For instance, we can conceive of search-based algorithms that generate outcomes but where the evaluation occurs through a trained neural network (Summerville et al., 2018).

3 THE CONTEXT OF AUTONOMOUS DESIGN TOOLS: A CONTROL PERSPECTIVE

The continuum from pure manual design to fully autonomous design highlights that autonomous design tools will operate in relation to multiple elements involved in the design process—human designers, autonomous design tools, and the environment. Fully autonomous design tools that define the design problem and devise solutions are a distant goal, and we need to consider these tools from a socio-technical perspective where human and machine designers interact synergistically. There are at least two reasons that require a human agent in such design systems. First, from an operational perspective, human designers *delegate* a design task to an autonomous tool, set parameters, start the autonomous design tool, and evaluate the outcome and make adjustments to the set of input parameters. Second, considering that problem spaces are evolving and that the same tool might be used for different design situations (and hence problem spaces), autonomous design tools suffer from the *frame problem*, which describes how algorithms are constrained by the rules (i.e., the knowledge) they currently possess and are hence incapable of reacting to environment states for which they are not prepared (Dennett, 2006; McCarthy & Hayes, 1981; Salovaara, Lyytinen, & Penttinen, 2019). In the case of symbolic approaches, the frame problem would demand that rules are added to the embedded design model to make it applicable to a broader or changing set of design problems (Dennett, 2006). However, even if we assume that we can infinitely add rules, such approach will increase the system's complexity and render its performance useless. While connectionist approaches involve learning, they still suffer from the frame problem as they are typically solving "closed-world" problems and remain constrained by the specific goal functions and available data (Salovaara et al., 2019).

Therefore, in this section, we turn to the interaction between autonomous design tools and their control units, most notably human designers, in relation to the environment in which they operate. We apply a control perspective (Mesarovic et al., 1970) to express the morphology of autonomous design tools. Figure 3, which is an extension of Figure 2, highlights the principal relationships of an autonomous design tool with its control system and environment.

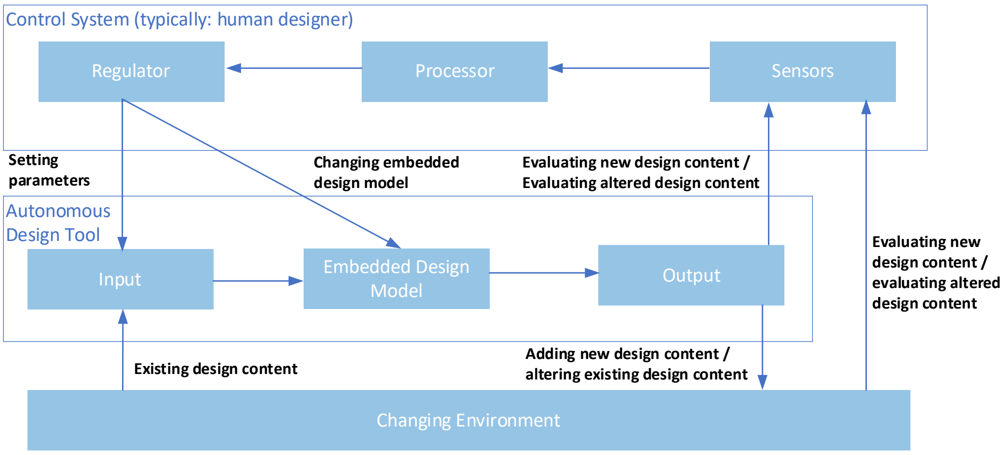

Control System (typically: human designer)
Regulator
Processor
Sensors
Setting parameters
Changing embedded design model
Evaluating new design content / Evaluating altered design content
Autonomous Design Tool
Input
Embedded Design Model
Output
Evaluating new design content / evaluating altered design content
Existing design content
Adding new design content / altering existing design content
Changing Environment

Figure 3. Control system, (partially) autonomous design tool, and the changing environment

The environment embodies the material and social context within which the design system as a whole operates and with which it interacts. It is likely to be changed through the use of the autonomous tool and its outcomes. The tool's input function refers to the interfaces through which the tool receives information about the environment. The embedded design model refers to the procedures followed to process information from the inputs, implemented through symbolic approaches, connectionist approaches, or a combination of such approaches. The output function represents the mechanisms through with the system effectuates the results of this process on the environment. An autonomous design tool is never entirely independent and its environment involves a second system—a control system—which triggers the autonomous design tool, monitors and evaluates its performance, and may even change the embedded design model in order to react to alterations in the problem space, thereby addressing the frame problem. On this view, model evolution can result from both the model's ability to learn and the intervention of the control unit changing the embedded design model (Seidel, Berente, Lindberg, et al., 2019). We describe the three components—control system, the autonomous design tool itself, and the environment—in what follows.

The human designer or design team, as a control system, involves three aspects: sensors, processors, and regulators. Sensors involve the designer's or design team's perception of the output of the autonomous tool. Of note is that while current applications typically involve human designers who work together with tools to create content (Seidel et al., 2018; Smelik et al., 2010a; Summerville et al., 2018) we can also think of non-human control systems and even autonomous design tools as control systems. However, as indicated earlier, addressing the frame problem will eventually require a human agent who is able to change the model to react to changes in the problem space. This creates a hierarchy of nested systems of human designers and autonomous design tools. This can involve the monitoring of performance measures applied to the design alternatives. Processing involves the interpretation and analysis of that sensory information to assess adequacy and the degree to which design preferences have been met. Regulators are the ways that designers change conditions of design activity. This could include modifying parameters or changing the design of the system as well as modifying or implementing a new algorithm.

The autonomous design tools receive sensory input through two channels: (1) through parameters specified by the designers programming or guiding the tool, that is, through the regulating component of the control system; (2) through input they receive from their interactions with the environment. Systems that are based on the symbol system approach, for instance, transform one symbol structure (e.g., an already existent representation of the design artifact) into another symbol structure (i.e., the new representation of the design artifact). The tool can make multiple design decisions without the intervention of the control unit. Eventually, however, some result will be evaluated by the control unit which may lead to new input and additional iterative cycles of deploying the tool. We describe the impact of an autonomous design tool on its environment in terms of the tool's output function. This design outcome might be a stand-alone artifact (e.g., a layout of a semi-conductor chip) or embedded artifact (e.g., modifications to a landscape in a video game, for instance, through adding a road network).

Finally, the environment is the context in which the autonomous tool operates and which it changes. The environment provides sensory inputs to the autonomous design tool. A search-based algorithm might receive a three-dimensional landscape as input and then generate alterations of this landscape until the process terminates with a satisfactory solution (Seidel et al., 2018). Similarly, a machine-learned model might be fed with a partial design and then complete that design (Summerville et al., 2018). Ultimately, whether or not the design outcome is satisfactory depends on how well it performs in the environment in which it is deployed. The environment can include both social (e.g., human stakeholders who have a say in whether an artifact meets the expectations) and technical (e.g., requirements of other components when designing more complex systems) elements. Table 2 provides an overview of how autonomous design tools, through their input and output functions, interact with control systems as well as the environment.

Table 2. Key components of autonomous design tools and their relationships to control systems and the environment

Component	Definition	Example
Input function (sensory)	Autonomous design tools receive input • from the designer guiding the autonomous tool (i.e., the regulating component of the control system) and • from the design environment.	The designer of a semiconductor chip sets parameters such as component parameters, physical parameters, and electrical parameters.
Embedded design model	The embedded design model—the algorithms and data models—determine how the tool designs; variants include: • models based on physical symbolic systems; • machine-learned models based using non-symbolic systems; • hybrids.	Can range from heuristics to machine-learned algorithms and may even involve a number of cooperating algorithms
Output function (actions)	The output function describes the actual actions that the tools takes with regards to its environment. The output function together with the embedded design model represent the actuating element of the autonomous tool as a goal-seeking system.	Autonomous design tools generate artifacts or change existing artifacts, for instance, the layout of a semiconductor chip.

Based on this conceptualization, we can further identify three key dimensions to characterize autonomous design tools: autonomy, interactivity, and understandability. First, the extent to which the autonomous tool requires pre-defined rules (either built-in or set by the control system such as a human designer) defines the level of autonomy. As indicated earlier, the less design tools depend on the prior knowledge of their designers (Russel & Norvig, 2016) the more autonomous they are.

Second, we can distinguish two types of interactivity: interactivity with the control system and interactivity with the environment. The more input is required from the regulator as part of the control system, the more interactive the design process is in terms of control-system-autonomous-tool interaction. Moreover, the autonomous tool may receive sensory input from the changing environment; the more input the tool receives from the environment that informs its course of action the more interactive the design process is in terms of tool-environment interaction.

Third, from the viewpoint of the human designer acting as the control unit, autonomous tools can exhibit different levels of understandability—while the functioning of a pathfinding algorithm for generating road networks may be relatively easily comprehended (and thus how the tool generated a result), this will be different in the case of neural networks training an artificial intelligence—reflecting recent discussions on explainability of artificial intelligence (Miller, 2018; Samek, Wiegand, & Müller, 2017).

Table 3 provides an overview these properties.

Table 3. Key properties of designer-autonomous-design-tools-systems

Property	**Description**	**Example**
Level of autonomy	Autonomous tools can rely on inputs provided by the designers (e.g., parameters) as well as information they receive from their interaction with the environment, i.e., with the problem space.	Chip design tools generate entire sections of a chip without direct intervention of the human designer.
Level of interactivity	While autonomous design tools can perform design activities with little to no user intervention, this does not mean that they operate in isolation. There are two types of interactivity: (1) Interactivity with the environment: the tool receives sensory input from the environment and acts upon this input, in turn changing the environment and generating new sensory input. (2) Interactivity with the control system: The tool receives input from the control system, be it a human designer or another tool.	The designer in the production of an asset (e.g., a landscape) for a video game monitors the process of the autonomous tool and, based on intermediate results, changes input parameters.
Level of understandability	The embedded design model of an autonomous design tool might be more or less easy to understand for the designer—or might be very complex.	Semiconductor chip designers cannot predict how the tool will layout components and also cannot always make sense of why particular design decisions were made by the tool.

4 ILLUSTRATIVE EXAMPLES: CONTENT GENERATION IN VIDEO GAMES

Autonomous design tools are now widely used to produce content for a new generation of video games. Current tools focus on procedural generation and mainly rely on symbolic approaches to identifying satisficing solutions. However, there are also some examples of learning algorithms, for instance: algorithms for terrain generation are trained on real-world terrains (Guérin et al., 2017). While tools make design decisions independently from the human designer, there is still significant interaction with human designers (Seidel, Berente, Lindberg, et al., 2019). Such algorithmically generated content can include a variety of game elements—including textures, buildings, road networks, etc. Designers typically combine these elements with specific hand-crafted elements. The interplay of automated and manual generation of content is crucial as humans are looking for rich and unique experiences, and undirected automated generation might lead to results that are not perceived as being authentic.

Ubisoft's *Ghost Recon Wildlands*, an action adventure game, is a recent example where designers used autonomous tools to generate large parts of the game space (Seidel, Berente, Lindberg, et al., 2019). Guided by human designers, algorithms procedurally generated much of the background content, and designers then tweaked what algorithms created and further handcrafted elements in the game space. In this process the tools would, for instance, generate large amounts of a detailed terrain. Then the designers would modify the terrain further and add extra detail. Some areas of the game space were still generated in a manual fashion. This combined process required developing and selecting appropriate tools and models that would align with the core concepts of the game as specified by a team of designers and developers. Next, we consider two examples from *Ghost Recon Wildlands* and interpret these examples through our conceptual lens.

The first example is the generation of a road network using a pathfinding algorithm.[2] The path finding algorithm transforms a data structure (a landscape without a road) into a different data structure (a landscape with a road). While the road itself is generated by the algorithm, this case is still characterized by interaction between the human designer—who acts as a control system—and the autonomous design tool. The human designer sets parameters (such as start and end points), runs the system, evaluates the outcome, and runs the tool again, until there is a satisfactory result. Importantly human designers are also involved in developing and selecting the specific algorithm and hence the design model embedded in the tool. Figure 4 highlights how different algorithms produce quite

[2] The process described here was inspired by Galin, Peytavie, Guérin, and Beneš (2011).

different designs. This illustrates how the selection of the algorithm—and hence the model embedded in the tool—is essential for the design outcome. Notably, this design outcome provides key input for further design steps which again involve the use of autonomous design tools, including for the generation of fences, crash barriers, traffic signs, road markings, specific types of grass or rocks on the roadside, powerlines along roads, etc. This indicates how the design outcomes generated by autonomous design tools fundamentally impact on the design process, including subsequent design decisions that both other autonomous design tools and human designers make.

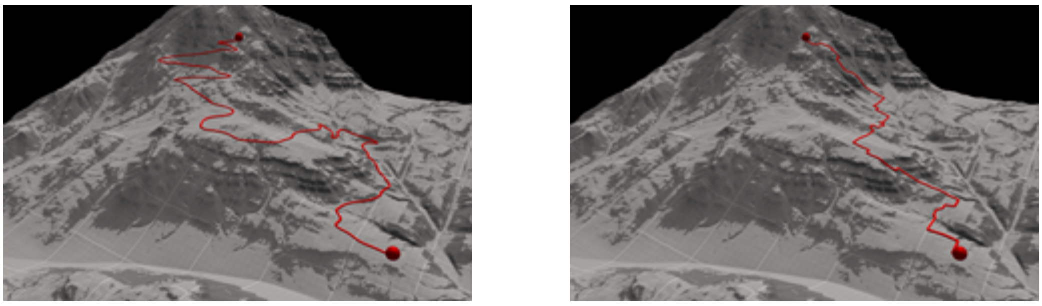


Figure 4. Generation of a road network using different algorithms (Source: Ubisoft)

In this example, all key components of an autonomous design tool and their context are present (Table 4). First, the tool receives sensory input (the topology of the map). Second, the tool computes a solution, in this case using a search-based algorithm, without much user intervention. Third, the tool acts upon the environment by adding the road network to the landscape, thereby altering the design artifact. Figure 5 shows an example of the output.

Figure 5. Autonomously generated road (Source: Ubisoft)

Table 4. Example: Generation of a road network in a videogame

Component		**Example**
Control system		The control system is a human designer using the tool to generate roads/a road network for a video game. The tool is executed by the designer and then evaluated by the human designer. The human designer must thus possess knowledge of the underlying embedded design model to anticipate what the algorithm does.
Autonomous design tool	Input function	Designer's specification in terms of start and end points Existing design content in terms of the landscape in which the road network is placed, e.g., the road can only have a certain incline otherwise an alternative path needs to be taken
	Embedded design model	Pathfinding algorithm The design tool searches the problem space by devising design alternatives
	Output function	Coordinates of the road network that fit to the landscape Alteration of design artifact, resulting in a landscape with road network
Environment		Altered design artifact: roads connecting start and end-points in the game space

Our second example, the generation of villages in the game space, allows us to further highlight how autonomous design tools and human designers interact (Figure 6).[3] This process starts with key decisions made by the human designer, including the identification of a center point for a village/town and the identification of related areas. These are key decisions that impact the road pattern within the town, and we can describe this process as a form of "architectural structuring" (Seidel, Berente, & Gibbs, 2019). The actual buildings are then placed by a self-aware packing algorithm. This process unfolds without human intervention and is based on building definitions, each of which has their own placement rules. Still, human designers have at their disposal tools to tweak what the algorithm has designed. The key is that placing the buildings involves design decisions that are made by a tool. Figure 6 displays the definition of a center for a village (I), the definition of the village boundaries (II), the definition of internal paths and zones (III), and the process of placing buildings (IV)—it is this stage where the design tool takes over.

[3] The process described here was inspired by Emilien, Bernhardt, Peytavie, Cani, and Galin (2012).

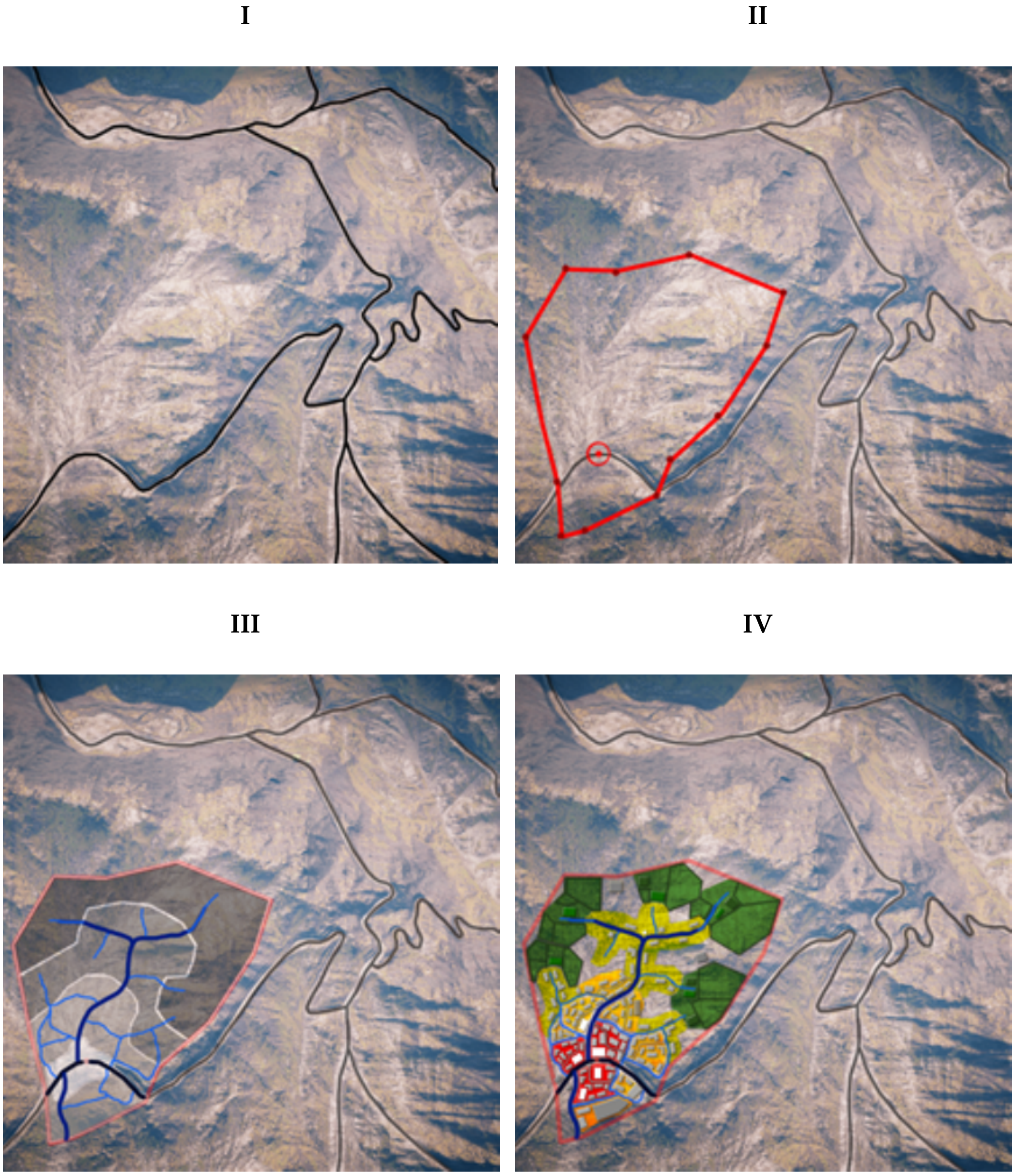


Figure 6. Steps in generating villages using a self-aware packing algorithm (Source: Ubisoft)

In this case all elements of autonomous design tools are present. The human designer acts as the control system and provides key inputs to the tool—such as the identification of areas to focus on. The autonomous design tool has an embedded design model in terms of a self-aware packing algorithm and the tool generates output that alters the environment in which the tool operates. Table 5 provides an overview and Figure 7 shows an example of a village generated using this approach.

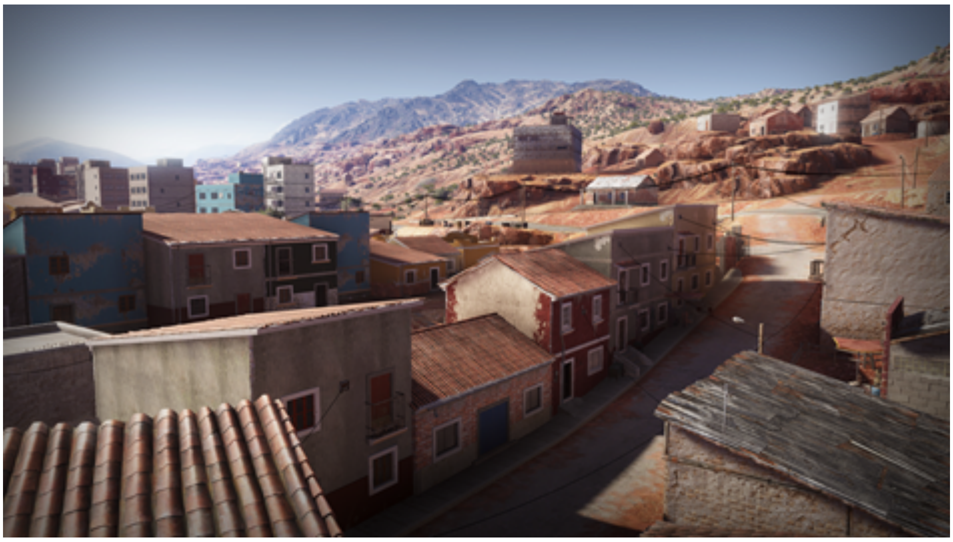


Figure 7. Village generated through interaction of human designer and autonomous design tools (Source: Ubisoft)

Table 5. Example: generation of villages in a videogame

Component		**Example**
Control system		The control system is a human designer who makes key architectural decisions: • Location of the village • Type of pattern (radial or square) • Internal road structure • Zoning • Decision on the buildings to use, including placement definitions for each building Each step is manually triggered so the user can visually validate the result before adjusting parameter of the current step or move to the next one. The human designer must thus possess knowledge of the underlying embedded design models to anticipate what the algorithms do.
Autonomous design tool	Input function	The center point of the village and boundaries A specific set of parameters for the different functions can be saved as a preset and reused elsewhere. Producing a different (for a different location) but predictable result in term of pattern and layout
	Embedded design model	Space partitioning Pathfinding Self-aware recursive packing algorithm
	Output function	Alteration of design artifact, resulting in a terraformed landscape with the village footprint

		Trajectories representing internal roads and paths that are used for further detailing through placing objects (lamppost, signs, etc.) on roadside 3d models of buildings
Environment		Altered design artifact: villages with dedicated center, roads, and other elements are added to the game space

With regards to the key properties of autonomy, interactivity, and understandability the two examples are comparable. First, the tools in these examples make design decisions on behalf of their designers, who however still have to provide user input. They can thus be described as being partially autonomous. There is interactivity as after setting parameters and running the tools again, the designers may still alter the resulting artifacts. This interactivity becomes particularly visible in the staged process of designing villages that moves from identifying a location to the actual placement of buildings and that involves interdependent designer and tool decisions. Finally, the embedded design models—such as the pathfinding algorithm and the self-aware packing algorithm—are quite understandable for the designers who use these tools.

5 DISCUSSION: A RESEARCH AGENDA FOR AUTONOMOUS DESIGN TOOLS AND CHANGING DESIGN WORK

Our key intention with this article is to provide a conceptual framework for studying the interactions between human and machine components in design systems that involve autonomous design tools, and therefore enabling theorizing of the materiality of autonomous design tools in relation to the organizing of design work. The literature on autonomous design tools (such as procedural generation) has so far largely focused on the technical aspects of implementing these approaches. Still, some scholars have indicated that these tools need to be considered in concert with the human designers employing such tools (Seidel, Berente, Lindberg, et al., 2019; Smelik et al., 2010b; Summerville et al., 2018). Hence, a socio-technical perspective on design tools becomes increasingly important as scholars have started to revisit expanded notions of material agency in the presence of increasingly autonomous and intelligent systems by using labels such as human-machine-learning (Seidel, Berente, Lindberg, et al., 2019), role-reversal (Demetis & Lee, 2017), digital agency (Ågerfalk, 2020), or meta-human systems (Lyytinen et al., 2020). Our conceptualization of autonomous design tools based on a rational agent perspective and control theory highlights how designing with autonomous tools is a process that is co-constituted by the activities of human designers and the design activities carried out by autonomous design tools. We have suggested that human designers act as control systems that "coach" autonomous design tools which act in a partially independent fashion in the sense that they make design decisions that cannot necessarily be anticipated by the human designers running the tools. However, despite increasing levels of system autonomy, humans still play a pivotal role as a control unit for the autonomous design tool.

The autonomous design tools we have discussed in this paper are tools designed for specific tasks. When Newell, Shaw, and Simon described their general problem solver (Newell, Shaw, & Simon, 1959), they conceived of a more general approach of computational problem solving based on the use of general heuristics of means-end-analysis and planning. Such a general approach to design still seems to be a distant goal. However, we have highlighted some developments in this direction such as using adversarial neural networks (Guérin et al., 2017) that foreshadow a development towards more flexible autonomous design tools. Following from this analysis of specificity and generality of tools, one key question is about the extent to which we can expect to find regularities in the way designers and machines interact when carrying out different tasks, and hence about the limits of theories about these new forms of human-machine interaction.

Against this background, autonomous design tools pose a variety of novel research challenges that recognize the socio-technical nature of designing with such tools. Here, we categorize these challenges into four areas to offer a systematic research agenda that can encourage interdisciplinary research teams to pursue fruitful and innovative research programs in this nascent field (Table 6).

Table 6. Research agenda

Phenomenon / level of analysis	**Example research questions**
Designer-autonomous-tool-interaction	How do humans and autonomous design tools interact *effectively* in design processes? How can the outcomes of using autonomous design tools be evaluated under different conditions; how to address the cognitive overload of human designers? How does learning take place when humans and autonomous tools interact? What forms of interaction and processes lead to better learning outcomes and design outcomes? How is such hybrid learning different from pure cognitive models of experiential learning or crafting? How do designers work with different types of embedded design models? What are the differences between interacting

	with design systems that are based on symbolic approaches versus those that are based on connectionist approaches?
Organizing design work with autonomous tools	How do autonomous design tools change designer roles, interactions, design principles, and organizing? How do key organizational processes such as decision making and sensemaking unfold in situations where human designers interact with autonomous design tools? How do organizational practices evolve as autonomous tools are introduced to design settings? Do organizational tasks or domains matter for how autonomous design tools are used and integrated?
Autonomous tools and markets/crowds/communities	How does the use of autonomous design tools change labor markets? Can autonomous design tools emerge as market-based agents that carry out specific design tasks and be offered as a service?
Ethical considerations of using autonomous design tools	What are the ethical dimensions and implications of using autonomous design tools? Are there regulatory issues related to recording and justifying design decisions and outcomes carried out by autonomous tools?

5.1 Designer-Autonomous Tool-Interaction

One important aspect that differentiates autonomous design tools from other types of software systems is that they generate outcomes where the human designer often cannot foresee the specifics of the outcome (Seidel, Berente, & Gibbs, 2019; Zhang et al., forthcoming). This is possible because these tools act autonomously as they move through the process of generating or altering an artifact while making invisible design decisions that do not depend on their designer’s (the one who designed the tool) prior knowledge of the design task as they go. Still, the designer makes initial assumptions about the design setting and goals (choosing tools, choosing parameters, setting parameters). This then yields contextual information (mostly about the design artifacts) which helps this designer to further guide the tool. Autonomous tools, through their independent design decisions, generate information that informs computations going forward, as well as the designer’s subsequent actions.

These observations indicate that we need to attend to the specific ways in which designers and tools engage with each other. It seems warranted to move our attention from the idea of designers enacting technology to processes of mutual enactment, where human activity and machine activity constitute each other in situ. However, as contemporary design tools such as those used in video game production still require designer input, we can ascribe a certain head status to the human designer (Leonardi, 2011). Still, we can conceive of a future where the boundaries of control and controlled will increasingly vanish, perhaps requiring a more symmetrical conceptualization of the relationship between control and controlled. There is no reason to believe that control in human-machine design systems could not reside in a machine or could be shared among human designers and autonomous design tools. The move from "technology enactment" to "mutual enactment" requires us to explore the specific ways designers interact with their tools and how they do so effectively. Moreover, we can expect that there will be new challenges in evaluating the outcomes generated by tools as well as through the interaction of designers and tools. Finally, it will be interesting to see how the nature of the embedded design model (symbolic versus connectionist or any combination) impacts on the interaction between human designers and tools.

5.2 Organizing Design Work with Autonomous Tools

It is likely that the increased use of autonomous design tools will involve moving away from an understanding of the designer as a craftsman (Sennett, 2008), towards being a tool chauffer. Designers increasingly need to develop a generalized understanding of the design problem as well as the envisioned solution so that they can think about appropriate strategies to generate design outcomes (which manifests in the selection and configuration of tools, including the selection of the embedded design model), instead of actively generating the design artifact through dedicated, manual design activities where each step is evaluated against the design goal. This requires us to rethink the way that we conceive of the institutional role of a designer, as it has potential implications for the way that education and learning will change across fields of practice.

In light of this changing role of the designer in relation to their materials and tools, it will further be important to explore if and how key organizational processes such as decision making related to participating or producing (March & Simon, 1993) as well as sensemaking (Weick, 1995; Weick, Sutcliffe, & Obstfeld, 2005) change in situations where autonomous tools become part of the fabric of organizing. Sensemaking, for instance, has been conceptualized as a retrospective process where not only cognition impacts action but where action impacts cognition (Weick, 2001)—but what does it mean for human cognition if this action is performed on their behalf by a machine with potentially unpredictable outcomes?

Finally, organizations want to understand the specific outcomes generated by autonomous design tools and how they fit into the overall product and service portfolio. While autonomous design tools promise to offload repetitive work from designers and quickly generate design artifacts of unprecedented scale with comparably little resources, it is also clear that these tools have limitations. While their output is complex and often unanticipated, they are still largely deterministic systems. The question arises to which extent these tools can indeed be creative in the sense that they generate truly novel artifacts; there is a risk that the tools will generate repetitive, perhaps boring (Backus, 2017) and non-creative content. Still, it seems reasonable to make two claims regarding the creativity involved while using these tools. First, if we conceive of autonomous design tools as part of a socio-technical design system where two components (humans and machines) interact, and where the output of each element impacts the action of the other, the overall system acts creatively, if it generates outcomes that are both novel and useful (Amabile, 1996) and that would not have been produced without such interactions. Second, the outcomes that are generated by the types of autonomous design tools we described in this paper exhibit a complexity that makes them unpredictable, and thus potentially novel, from the designers' point of view (Seidel, Berente, & Gibbs, 2019).

5.3 Autonomous Tools and Markets/Crowds/Communities

The described changes in the way humans and machines interact as well as in the way we organize for work can be seen as micro foundations for broader level changes at multiple levels of analysis. We may, for instance, expect that the labor market will, going forward, require different designer skills. Specifically, designers will require in-depth knowledge about how to select, orchestrate, and run autonomous design tools. Moreover, software development skills will be important for designers as they seek to understand and perhaps alter the models embedded in autonomous design tools.

Moreover, it will be interesting to see to what extent autonomous design tools will not only be used to create products, but also function as market-based agents that offer services. In the past, software-as-a-service and related concepts have mainly focused on providing capabilities such as for data storage and process automation. If autonomous design tools become market-based agents that carry out design tasks on behalf of a customer, organizations will rely on external stakeholders to perform design work. This bears the potential for disrupting a variety of industries, as the generation and implementation of purposeful design outcomes is a key source of value generation in many contemporary organizations. What, however, would the consequences be if such tasks could be performed at higher speed, higher scale, and perhaps decreased cost by an external provider?

5.4 Ethical Considerations of Using Autonomous Design Tools

Finally, we have to attend to the ethical dimensions and implications of using autonomous design tools. For instance, these tools deeply penetrate into the types of work that have traditionally be seen to be reserved for humans—work that is related to creativity and design. We can thus expect that these tools will challenge established role identities of designers and related professions and that may even lead to situations where designers feel threatened by that technology (Seeber et al., 2020). Following from this observation, it is crucial to explore the pertinent regulatory issues related to recording and justifying design decisions and outcomes carried out by autonomous tools. This involves questions with regards to the intellectual property that is generated by autonomous design tools as well as the consequences of using such intellectual property.

6 CONCLUSION

In this paper we have discussed the conceptual foundations of autonomous design tools. These foundations prepare the ground to study how these tools are involved in socio-technical systems and how they change how we organize design work. To this end, we have highlighted how designers currently have tools at their disposal that range from tools which provide limited support for manual tasks, to design tools which are fully autonomous. Moreover, we have argued that the idea of fully autonomous design tools remains an abstraction; the practical examples we have identified in areas such as the design of video games, which formed the baseline example in this paper, rely on the interaction of human designers and tools. We also distinguished two general approaches to building autonomous design tools (physical symbol systems and connectionist systems) and we have highlighted how there is now a nascent interest in tools that learn from interactions with their environment, thus moving us closer to the vision of fully autonomous design tools.

After having experienced two AI winters, AI and associated design systems are finally flourishing. These developments have been driven by vast amounts of available data upon which machine learning algorithms are capitalizing, as well as the emergence of cloud-based computing infrastructures that provide the necessary fuel, the computing power necessary to explore vast design spaces. The emergence of these technologies heralds a possible revolution in how we think about design across multiple domains. It is therefore incumbent on us to seek to thoroughly understand this new breed of tools and the consequences of their usage.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under grants 2026583, 1909803, 1717473, and 1745463.

REFERENCES

Ågerfalk, P. J. (2020). Artificial intelligence as digital agency. *European Journal of Information Systems*, 29(1), 1-8.

Amabile, T. M. (1996). *Creativity in context*: Westview Press.

Ashlock, D., & McGuinness, C. (2013). Landscape automata for search based procedural content generation. In *2013 IEEE Conference on Computational Intelligence in Games (CIG).*

Backus, K. (2017). Managing output: Boredom versus chaos. In T. X. Short & T. Adams (Eds.), *Procedural Generation in Game Design* (pp. 13-21): AK Peters/CRC Press.

Badue, C., Guidolini, R., Carneiro, R. V., Azevedo, P., Cardoso, V. B., Forechi, A., . . . Mutz, F. (2019). Self-driving cars: A survey. arXiv preprint arXiv:1901.04407.

Baskerville, R., Myers, M., & Yoo, Y. (2019). Digital first: The ontological reversal and new challenges for IS. *MIS Quarterly, 44(2), 509-523.*

Baxter, R. J., & Berente, N. (2010). The process of embedding new information technology artifacts into innovative design practices. *Information and Organization*, 20(3-4), 133-155.

Boden, M. A. (2009). Computer models of creativity. *AI Magazine*, 30(3), 23-34.

Boland, R. (2004). Design in the punctuation of management action. In R. Boland & F. Collopy (Eds.), *Managing as designing: Creating a vocabulary for management education and research* (106-112). Stanford, California: Stanford Business Books

Brown, C., & Linden, G. (2011). *Chips and change: How crisis reshapes the semiconductor industry*: MIT Press.

Buchanan, R. (1992). Wicked problems in design thinking. *Design Issues*, 8(2), 5-21.

Cassell, J., Sullivan, J., Churchill, E., & Prevost, S. (2000). *Embodied conversational agents*: MIT press.

Chang, T.-C., & Wysk, R. A. (1997). *Computer-aided manufacturing*: Prentice Hall.

Daugherty, P. R., & Wilson, H. J. (2018). *Human + machine: Reimagining work in the age of AI*: Harvard Business Press.

Demetis, D., & Lee, A. (2017). When humans using the IT artifact becomes IT using the human artifact. In *50th Hawaii International Conference on System Sciences.*

Dennett, D. C. (2006). Cognitive wheels: The frame problem of AI. In C. Hookway (Ed.). *Minds, Machines and Evolution* (pp. 129-150): Cambridge University Press.

Dorst, K., & Cross, N. (2001). Creativity in the design process: Co-evolution of problem-solution. *Design Studies*, 22(5), 425–437.

Emilien, A., Bernhardt, A., Peytavie, A., Cani, M.-P., & Galin, E. (2012). Procedural generation of villages on arbitrary terrains. *The Visual Computer*, 28(6-8), 809-818.

Galin, E., Peytavie, A., Guérin, E., & Beneš, B. (2011). Authoring hierarchical road networks. In *Computer Graphics Forum* (Vol. 30, No. 7, pp. 2021-2030). Oxford, UK: Blackwell Publishing Ltd.

Gero, J. S. (1990). Design prototypes: a knowledge representation schema for design. *AI Magazine*, 11(4), 26-36.

Goel, V., & Pirolli, P. (1992). The structure of design problem spaces. Cognitive science, 16(3), 395-429.

Guérin, É., Digne, J., Galin, E., Peytavie, A., Wolf, C., Benes, B., & Martinez, B. (2017). Interactive example-based terrain authoring with conditional generative adversarial networks. *ACM Transactions on Graphics (TOG)*, 36(6), 228.

Gupta, K. C., Garg, R., & Chadha, R. (1981). *Computer aided design of microwave circuits*. NASA STI/Recon Technical Report A, 82.

Hendrikx, M., Meijer, S., Van Der Velden, J., & Iosup, A. (2013). Procedural content generation for games: A survey. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 9(1), 1:2-1:24.

Krish, S. (2011). A practical generative design method. *Computer-Aided Design*, 43(1), 88-100.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*: Oxford University Press.

Leonardi, P. M. (2011). When flexible routines meet flexible technologies: Affordance, constraint, and the imbrication of human and material agencies. *MIS Quarterly*, 35(1), 147-168.

Liapis, A., Yannakakis, G. N., & Togelius, J. (2014). Computational game creativity. In Proceedings of the *5th International Conference on Computational Creativity*.

Lyytinen, K. a., Nickerson, J. V., & King, J. L. (2020). Metahuman systems = humans + machines that learn. *Journal of Information Technology*.

March, J., & Simon, H. A. (1993). *Organizations* (2nd ed.). New York: Wiley.

Markus, M. L., & Silver, M. S. (2008). A foundation for the study of IT effects: A new look at DeSanctis and Poole's concepts of structural features and spirit. *Journal of the Association for Information Systems*, 9(10), 609-632.

McCarthy, J., & Hayes, P. J. (1981). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer & D. Michie (Eds.), *Machine Intelligence 4* (pp. 463-502). Edinburgh: Edinburgh University Press.

Mesarovic, M. D., Macko, D., & Takahara, Y. (1970). *Theory of hierarchical, multilevel, systems* (Vol. 68). New York: Academic Press.

Miller, T. (2018). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1-38.

Müller, P., Wonka, P., Haegler, S., Ulmer, A., & Van Gool, L. (2006). Procedural modeling of buildings. *ACM Transactions on Graphics (TOG)*, 25(3), 614-623.

Newell, A., Shaw, J. C., & Simon, H. A. (1959). Report on a general problem-solving program. In *IFIP Congress*.

Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104, No. 9): Prentice-Hall Englewood Cliffs, NJ.

Orlikowski, W. J. (2000). Using technology and constituting structures: A practice lens for studying technology in organizations. *Organization Science*, 11(4), 404-428.

Parish, Y. I., & Müller, P. (2001). Procedural modeling of cities. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*.

Parker, G., Van Alstyne, M., & Jiang, X. (2016). *Platform ecosystems: How developers invert the firm*. Boston University Questrom School of Business Research Paper.

Pickering, A. (1993). The mangle of practice: Agency and emergence in the sociology of science. *American Journal of Sociology*, 99(3), 559-589.

Pickering, A. (1995). *The mangle of practice: Time, agency and science*. Chicago, IL: The University of Chicago Press.

Pohl, I. (1970). Heuristic search viewed as path finding in a graph. *Artificial Intelligence*, 1(3-4), 193-204.

Porter, M. E., & Heppelmann, J. E. (2014). How smart, connected products are transforming competition. *Harvard Business Review*, 92(11), 64-88.

Puranam, P., Alexy, O., & Reitzig, M. (2014). What's "new" about new forms of organizing? *Academy of Management Review*, 39(2), 162-180.

Rai, A., Constantinides, P., & Sarker, S. (2019). Editor's comments: Next-generation digital platforms: Toward human–AI hybrids. *MIS Quarterly*, 43(1), iii-ix.

Russel, S., & Norvig, P. (2016). *Aritifical intelligence. A modern approach*: Pearson.

Salovaara, A., Lyytinen, K., & Penttinen, E. (2019). High reliability in digital organizing: Mindlessness, the frame problem, and digital operations. *MIS Quarterly*, 43(2), 555-578.

Samek, W., Wiegand, T., & Müller, K.-R. (2017). *Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models*. arXiv preprint arXiv:1708.08296.

Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G.-J., Elkins, A., . . . Randrup, N. (2019). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, 103174.

Seeber, I., Waizenegger, L., Seidel, S., Morana, S., Benbasat, I., & Lowry, P. B. (2020). Collaborating with technology-based autonomous agents. *Internet Research*, 30(1), 1-18.

Seidel, S. and Berente, N. (2020) "Automate, Informate, and generate: Affordance primitives of smart devices and the Internet of Things," in S. Nambisan, K. Lyytinen, & Y. Yoo (Eds.), *Handbook of Digital Innovation*, Northampton: Edward Elgar Publishing.

Seidel, S., Berente, N., & Gibbs, J. (2019). Designing with autonomous tools: Video games, procedural generation, and creativity. In *Proceedings of the 40th International Conference on Information Systems*.

Seidel, S., Berente, N., Lindberg, A., Nickerson, J. V., & Lyytinen, K. (2019). Autonomous tools & design work: A triple-loop approach to human-machine learning. *Communications of the ACM*, 62(1), 50-57.

Seidel, S., Berente, N., Martinez, B., Lindberg, A., Lyytinen, K., & Nickerson, J. V. (2018). Succeeding with autonomous tools in systems design: Reflective Practice & Ubisoft's Ghost Recon Wildlands Project. *IEEE Computer*, 51(10), 16-23.

Sennett, R. (2008). *The craftsman*. London: Allen Lane.

Shaker, N., Togelius, J., & Nelson, M. J. (2016). *Procedural content generation in games*: Springer.

Simon, H. A. (1996). *The sciences of the artificial*. Cambridge, MA: MIT Press.

Smelik, R. M., Tutenel, T., de Kraker, K. J., & Bidarra, R. (2010a). Integrating procedural generation and manual editing of virtual worlds. In *Proceedings of the 2010 Workshop on Procedural Content Generation in Games*.

Smelik, R. M., Tutenel, T., de Kraker, K. J., & Bidarra, R. (2010b). Interactive creation of virtual worlds using procedural sketching. In *Eurographics (Short papers)* (pp. 29-32).

Smith, A. M., & Mateas, M. (2011). Answer set programming for procedural content generation: A design space approach. *IEEE Transactions on Computational Intelligence and AI in Games*, 3(3), 187-200.

Smolensky, P. (1987). Connectionist AI, symbolic AI, and the brain. *Artificial Intelligence Review*, 1(2), 95-109.

Summerville, A., Snodgrass, S., Guzdial, M., Holmgård, C., Hoover, A. K., Isaksen, A., . . . Togelius, J. (2018). Procedural content generation via machine learning (PCGML). *IEEE Transactions on Games*, 10(3), 257-270.

Sun, R. (1999). Artificial intelligence: Connectionist and symbolic approaches. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences*, 783-789: Elsevier.

Tiwana, A. (2015). Evolutionary competition in platform ecosystems. *Information Systems Research*, 26(2), 266-281.

Togelius, J., Yannakakis, G. N., Stanley, K. O., & Browne, C. (2011). Search-based procedural content generation: A taxonomy and survey. *IEEE Transactions on Computational Intelligence and AI in Games*, 3(3), 172-186.

Veale, T., Cardoso, F. A., & y Pérez, R. P. (2019). Systematizing creativity: A computational view. In T. Veale & A. Cardoso (Eds.), *Computational Creativity* (pp. 1-19): Springer.

Weick, K. E. (1995). *Sensemaking in organizations* (Vol. 3): Sage.

Weick, K. E. (2001). *Making sense of the organization*. Malden, MA, USA: Blackwell Publishing.

Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the process of sensemaking. *Organization Science*, 16(4), 409-421.

Yumer, M. E., Asente, P., Mech, R., & Kara, L. B. (2015). Procedural modeling using autoencoder networks. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*.

Zammuto, R. F., Griffith, T. L., Majchrzak, A., Dougherty, D. J., & Faraj, S. (2007). Information technology and the changing fabric of organization. *Organization Science*, 18(5), 749-762.

Zhang, Z., Yoo, Y., Lyytinen, K., & Lindberg, A. (forthcoming). The unknowability of autonomous tools and the liminal experience of their use. *Information Systems Research*.

Zuboff, S. (1988). *In the age of the smart machine: The future of work and power* (Vol. 186): Basic books New York.