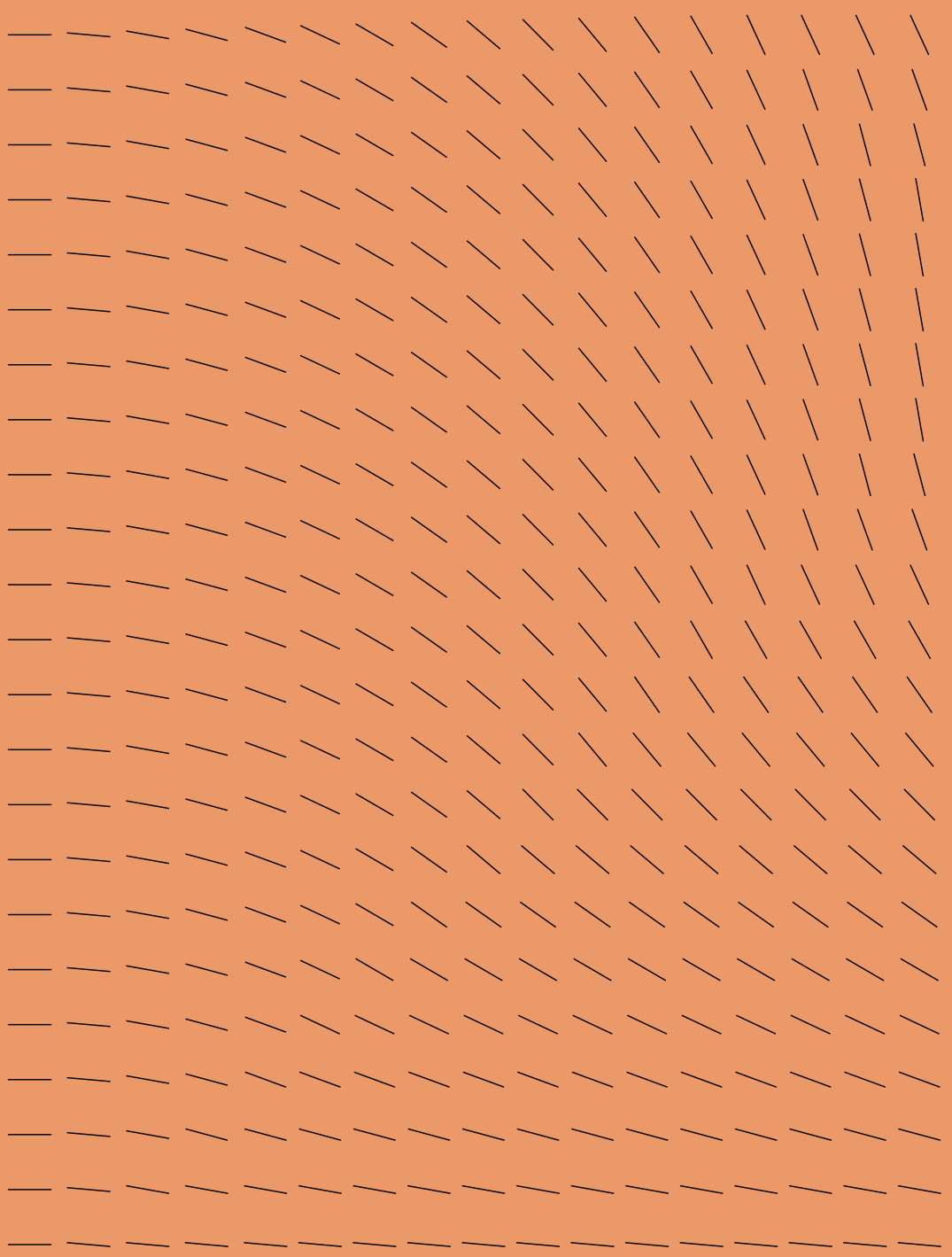
weizenbaum
institut



Proceedings of the Weizenbaum Conference

Practicing Sovereignty. Interventions for Open Digital Futures

June 2022

Weizenbaum Institute for the Networked Society –
The German Internet Institute

Proceedings of the Weizenbaum Conference 2022: Practicing Sovereignty. Interventions for Open Digital Futures

Proceedings of the Weizenbaum Conference 2022: Practicing Sovereignty. Interventions for Open Digital Futures

Berlin, 2023
DOI 10.34669/wi.cp/4 \ ISSN 2510-7666

Weizenbaum Institute for the Networked Society –
The German Internet Institute
Hardenbergstraße 32 \ 10623 Berlin \ Tel.: +49 30 700141-001
info@weizenbaum-institut.de \ www.weizenbaum-institut.de

VOLUME EDITORS:
Bianca Herlo
Daniel Irrgang

EDITORIAL MANAGER:
Moritz Buchner

WEIZENBAUM INSTITUTE: The Weizenbaum Institute for the Networked Society – The German Internet Institute is a joint project funded by the Federal Ministry of Education and Research (BMBF) and the State of Berlin. It conducts interdisciplinary and basic research into the digital transformation of society through digitization and provides politicians, business and civil society with evidence- and value-based options for action in order to shape digitization in a sustainable, self-determined and responsible manner.

This work has been funded by the Federal Ministry of Education and Research of Germany (BMBF) (grant no.: 16DII121, 16DII122, 16DII123, 16DII124, 16DII125, 16DII126, 16DII127, 16DII128 – "Deutsches Internet-Institut").

Table of Contents

Proceedings of the Weizenbaum Conference 2022:
Practicing Sovereignty. Interventions for Open Digital Futures

DIGITALLY AIDED SOVEREIGNTY

A SUITABLE GUIDE FOR THE E-GOVERNMENT TRANSFORMATION?

Herzog, Christian
Ethical Innovation Hub, Institute for Electrical Engineering in Medicine, University of Lübeck,
Lübeck, Germany
christian.herzog@uni-luebeck.de

Zetti, Daniela
Department of Science, Technology and Society, Technical University of Munich
Munich, Germany
daniela.zetti@tum.de

KEYWORDS

digital sovereignty; e-government; digitally aided sovereignty; e-democracy; participation

DOI: 10.34669/wi.cp/4.1

ABSTRACT

We advocate for the adoption of an integrated strategy aimed at achieving increased participation *via* effective digital public administration services. We argue that it is urgent to understand the integration of participatory approaches from the field of e-democracy in digitalized public administration, as trendsetting e-government implementations are already underway. We base our arguments on the observation that the approaches in e-democracy and e-government seem to be locked into extremes: In e-democracy, (experimental) platforms have failed to create a participative political culture. E-government, in turn, narrowly perceives citizens as customers. Additionally, efforts to increase digital sovereignty have mostly been educational ones that support citizens' self-determined use *of the digital* but do not address sovereignty *via the digital*. As a result, digitalized public administration is not achieving its potential to create opportunities for participation during encounters with the administration. Hence, we argue for the adoption of a *digitally aided sovereignty* as a normative guide for an e-government transformation that strives to create opportunities for participation *via the digital*.

1 INTRODUCTION

The term digital sovereignty is perhaps most commonly subdivided into types referring to the state's, the economy's, or the individual's control over the digital (Floridi, 2020; Moerel & Timmers, 2021). Based on their ethnographic work, Couture and Toupin (2019) have disaggregated perspectives on sovereignty "when referring to the digital" even further. They have identified (i) *digital sovereignty of governments and states*, as a nation's control over digital infrastructures, (ii) *cyberspace sovereignty*, as the notion that globalized networks transcend state sovereignty, (iii) *indigenous digital sovereignty*, as a notion from an indigenous perspective regarding control over cultural data, (iv) *the digital sovereignty of social movements*, as a contrasting notion to the *digital sovereignty of governments and states*, aimed at creating viable alternatives to commercial or state-sponsored digital infrastructures, and (v) *personal digital sovereignty* as a notion that "refers to the control of an individual over their data, device, software, hardware, and other technologies" (Couture & Toupin, 2019). This contribution mostly focuses on the fourth and fifth conception; hence, we would like to advocate for stronger consideration of individual or citizen sovereignty, referring to the notion that in democratic states, "the people is the sovereign" (Merkel, 2020).

Consequently, we do not ask what it takes to have sovereignty over the digital—that is, infrastructures, tools, technologies, and data—but rather look at how the digital may help individuals to exert sovereignty. To the best of our knowledge, most of the discussion surrounding digital sovereignty deals with control *over* the digital, while much less has been said about how to exert sovereignty *via* the digital. Granted, the latter may not be what is commonly associated with the term *digital sovereignty*. As a solution, we propose to use the term *digitally aided sovereignty* when referring to the digital as a facilitator of a citizen's or person's legitimate authority in a democratic state. However, the two notions seem to be interrelated to some degree, because often sovereignty *over* the digital is, in fact, a prerequisite for the realization of sovereignty *via* the digital. The point is not to criticize the term "digital sovereignty"[1] but to offer a fresh and argumentative perspective on what may guide good e-government solutions.

We will proceed by discussing how efforts in e-democracy and e-government may be failing to realize participatory value and thereby strengthening administrative and governmental accountability. We will then argue that adopting *digitally aided sovereignty* as a guide for the e-government transformation would be more clearly aligned with these goals. We conclude by outlining potential directions to take.

[1] Others have done a far more proficient job at this; for instance, (Moerel & Timmers, 2021; Pohle & Thiel, 2020, 2021).

2 E-DEMOCRACY AND E-GOVERNMENT: QUO VADIS?

The *Memorandum on E-Government* defines e-government as "the execution of processes of public opinion formation, decision-making, and performance of functions in politics, state, and administration with the intensive use of information technology" (GI and VDE 2000). To a large extent, this definition encompasses e-democracy, which is about "citizen participation through [information and communications technology] to support legitimate representation of citizens in a democratic society" (Christiansen, 2010). However, we distinguish between e-government and e-democracy, because, in practice, the term e-government is arguably less concerned with participation or in aiding individuals in living out their own sovereignty via electronic services (Grönlund, 2010) than would be necessary for the purposes of this contribution. For instance, while the definition of e-government presented by GI and VDE is very broad, the OECD has categorized definitions into four types, of which only one includes a notion of improving government; none of these contains an explicit reference to participation. From this starting point, we will attempt a brief characterization of current developments.[2]

2.1 E-GOVERNMENT IS TRAPPED IN THE PROVIDER/CONSUMER PERSPECTIVE

Germany has planned to have digitally transformed all 575 of its public services by 2022, which, according to Mergel (2021), is a more ambitious agenda than is evident in any other country. While there may be more than just a few delays, this figure points to the urgent need to critically assessed the directions taken. After all, inertia—resulting from both human agents as well as organizational, political and technological factors—is not only impeding or delaying the initial transformation *towards* the digital (see, e.g., Friesike & Sprondel, 2022; Mokyr, 2000; Schmid et al., 2017) but may also slow down or prevent potentially necessary corrections *after* its deployment.

Hence, a look at the report on implementing digitalization in Germany (*Digitalisierung gestalten – Umsetzungsstrategie der Bundesregierung*, 2020) may shed some light on the current situation. On the subject of the "modern state," the report lists two foci: the state as a service provider and the digitalization of public administration. The former implies a business-like conceptualization of the relationship between the citizen and the state. Indeed, the "Digital Single Market Strategy for Europe" (European Commission, 2015) claims that it is "crucial to increasing cost-efficiency and quality of the services provided to citizens and companies" and proposes the "once only" principle (Pernice, 2016). The "once only" principle essentially refers to secure ways to reuse data provided by

[2] We will focus on developments in Germany. However, we believe that most may hold true in other countries as well.

citizens, effectively reducing the amount of contact required between citizens and public agencies. Indeed, the projects listed in Germany's digitalization report include the establishment of an e-payment processing platform for administrative services, online portals for health information, or digitalized backends that connect different administrative agencies.

While there is much to be said for increasing transparency or making it easier for citizens to (digitally) contact the public administration, we concur with Pohle and Thiel (2020), who write that in "many instances, citizens are being reduced to consumers of digital services rather than valued in their capacity as democratic citizens." Bekkers and Zouridis (1999) go even further and argue that there is a risk of destroying active political engagement by citizens if they are only viewed as consumers. It seems clear that the transformation towards digital public administration does not aim to create increasingly participative processes. There are thus grounds to hope that, unlike in the UK almost two decades ago (Hazlett and Hill 2003), moves toward e-government will not just exacerbate the shortcomings of offline public administration in the online world. However, successful endeavors in the fields of e-participation and e-democracy—in parallel to the digitalization of public administrative services—will determine whether citizens are really just viewed as consumers. After laying out how approaches in e-democracy seem to be few and far between and how they are failing to drive public engagement, we will continue making a case for integrating participation into digital public administration.

2.2 E-Democracy is Failing to Engender Increased Participation

Disappointingly, with regard to its second focus, the report "Digitalisierung gestalten" (2020) only lists a single project that refers to "digital participation and forms of online dialog." All other projects are either educational, informational, or internal to administrative agencies. On closer inspection, the "online dialog" turns out to be the respective federal ministry's use of social media and the use of communication via messenger apps. Hence, little seems to have changed in the years since Winkel (2007, p. 14) attested that in Germany "applications enabling result-oriented participation of citizens in political decision-making processes are encountered only as rare exceptions, even on the local level." A literature review commissioned by the European Parliament's Science and Technology Options Assessment Panel (STOA) found that e-democracy has been developing at a much slower pace than e-government (European Parliament. Directorate General for Parliamentary Research Services (DGPRS). 2017a, p. 56). The study comes to disillusioning conclusions, stating that "it appears that, at times, projects that at first glance appear to be participative turn out not to be consultative or deliberative in nature, but have the objective of informing citizens about decisions that have already been made" (European Parliament. DGPRS. 2017b, p. 9). One of the reasons

mentioned for this is the "often experimental character" (European Parliament. DGPRS. 2017b, p. 11).

According to Chadwick and May (2003) and later confirmed in a review by Madsen, Berger, and Phythian (2014), the lack of effort in advancing e-democracy can be attributed to the widespread adoption of a "managerialism" stance, which mainly focuses on the efficient delivery of information and services and rests on the assumption that the (rapid) provision of information equates to having an open government. We suggest that this may still hold true today and that this mindset lacks an emphasis on advancing consultative or participative practices, which may further explain the gap between e-democracy and e-government.

A complete analysis of why e-democracy and successful formats for e-participation have failed to prevail is beyond the scope of this contribution, and many ideas can be found in the European Parliament's "Prospects for E-democracy" study summary (European Parliament. DGPRS. 2017b). Certainly, issues in hardware accessibility, digital and administrative literacy gaps, and socio-economic hurdles have prompted slow adoption of both e-government and e-democracy (Mergel, 2021). As Winkel (2007, p. 8) writes, "[w]hosoever wants more participation, must consider that participation procedures can cause high expenditures of time and money." Of course, this extends to those participating. There are, of course, impressive examples and prototypes. For instance, Herlo, Stark, and Bergmann (2021) have devised a hybrid (real-world/virtual) artifact for "more inclusive modes of participation in urban development projects".

However, participation has not yet been integrated on a systemic level—for instance, it has not been included in the practices of digitalized public administration. In the words of a report of the Scientific Foresight Unit (European Parliament. DGPRS. 2017b, p. 11), "the lack of any impact on decision making is one [of] the most striking findings." There is work on integrating the citizen perspective when designing online public services (Mergel, 2021, p. 343), but ways of using e-government for empowerment by offering services that facilitate each individual's capacity to act and participate as a democratic citizen have yet to manifest.

3 From Digital Sovereignty to Digitally Aided Sovereignty—A Guide for the e-Government Transformation?

In practice, the term "digital sovereignty" seems to imply a strategic focus on building competencies and increasing knowledge regarding the digital. Funding programs have also focused on creating new ways of interaction in the name of digital sovereignty. For instance, Germany's Federal Ministry of Education and Research issued a call for proposals in 2019 for "Human Technology Interaction for

Digital Sovereignty" with the goal to "promote the development of new digital forms of interaction and human-technology dialogs conducive to learning for the reflective handling of data and digital technologies." (Germany's Federal Ministry of Education and Research 2019, own translation). While the ultimate goal has been to increase the sovereignty of citizens to handle their data, the means seem to be narrowly focused on educational approaches. This indicates a rationale that seems to equate enhanced knowledge about data flows, protection and usage with increased control over the digital—not unlike Chadwick and May's (2003) assessment of the "managerialism" approach to e-government. The selection of research projects that received funding seems to confirm this: They include approaches that seek to improve the capacities "of adolescents to secure their data via micro games" and strengthen the ability of "less technologically affine people […] to use electronic health records" as well as approaches to devise "data visualizations [that] should allow users to sovereignly decide which products to use" or "competency-enhancing teaching and learning scenarios" ("Mensch-Technik-Interaktion für Digitale Souveränität" 2020, own translations). These projects envision little in terms of allowing citizens to exert actual control over data, since they limit themselves to developing informational and educational systems, which appear to be mainly geared towards the acceptance of novel digital practices rather than empowering citizens.

However, if we conceive of digital sovereignty as an extension of legitimate power (sovereignty) into the digital realm, this digital realm might be conducive to enabling citizens to exert their rightful sovereignty itself instead of only exerting sovereignty over the digital. With regard to e-government, this demands a critical assessment of how the potential of digital platforms and tools can be harnessed to increase opportunities to participate in an inclusive manner.

In the vein of such *digitally aided sovereignty*, we would like to propose that, at least in some ways, e-democracy and e-participation should be thought of as integrated into e-government services. This would refocus e-government from relying on citizens' sovereignty *over* the digital to being driven by aiding citizens' sovereignty *via* the digital. The latter does not devalue the need for the former. On the contrary, digitally aided sovereignty may well be seen as subsidiary to digital sovereignty. However, a shift in emphasis when conceptualizing e-government practices would clearly underscore that the goal is firmly set on increasing citizens' sovereignty.

Authors such as Pratchett (2006) make a case that e-democracy and e-government have distinct modes of operation and should be approached very differently despite their similarities. On some level, we concur. Some public services should not be bloated due to requests for participation but should instead be precisely framed and to the point. However, given the hesitant adoption of e-democracy approaches and the lack of a clear impact, it stands to reason that encounters with public administrations may be about more than just executing a specific administrative task. The next section

sketches out a few ideas and possibilities for integrating participatory elements into e-government systems following the idea of digitally aided sovereignty.

4 POTENTIAL DIRECTIONS TO TAKE

Even though this contribution can only provide a cursory sketch, in the following, we'd like to outline what it could mean to take the idea that e-government practices should be guided by digitally aided sovereignty seriously, a notion that would hence extend to the digitalization of public administration services. It proceeds from Pernice's (2016) observation that "strengthening the relationship between the citizens and their political institutions and leaders, be it at local, at regional, at national or at the European level has yet to be explored."

- **Integrating participatory and consultative elements into specific digitalized public administration services:** Clearly, by envisioning encounters with digital public administration within virtual or hybrid settings, digitally aided platforms must extend beyond the educational. Visualizations of data flows (Stowers, 2013) or electronic forms could provide the option of leaving item specific feedback as a means of implementing a more consultative approach. This could address the complaint that the topic on which an e-consultation is being run is too broad (European Parliament. DGPRS. 2017b, p. 9). Possibilities for consultations that are placed right at the point where a specific associated service is used may alleviate this issue.
- **Integrating decision elements into educational approaches:** Digitally aided sovereignty could be expanded by simply extending educational approaches currently developed to increase digital competencies. For instance, tools for exploring and visualizing data flows between public agencies could integrate simple interfaces that would allow users to refuse specific data transfers in accordance with citizens' rights. Even though such designs may go against the "once only" doctrine, citizens could even be asked to explicitly allow certain data transfers within interfaces that visualize flows and even highlight implications. Instead of hiding away what public administration agencies do for citizens, public agencies may elicit more impactful engagement by actually requiring citizens to understand and act upon certain administrative duties and workflows. This would require a delicate balance to be struck, as increased demands on effort and time may lead to undue hardships for the marginalized. A framework of tolerant paternalism (Floridi, 2015) that both boosts citizen's information level and prompts explicit choices for their benefit could guide implementation.
- **Devising public administration services as encounters from the perspective of experience design:** Folk wisdom has it that administrative duties and interactions with public

administration itself are off-putting, tedious, and often experienced in terms of "them against us," rather than in a sustained atmosphere of cooperation. However, there may be ways to design encounters with administrations within a framework of mutual responsibility that helps citizens to exert their sovereignty. Again, this may possibly mean loosening up on the "once only" dogma and specifically promoting additional virtual, real, or hybrid encounters with public administrations. If supported by digital tools that are easy to use, unintrusive, but effective, administrative duties can become the basis for respectful encounters and perhaps increase the willingness to participate. The design of interactions must make this tangible and take up the "[…] responsibility of a democratic government to help furnish whatever services and resources are needed to prepare citizens for active, effective, and intelligent engagement", (Schuler, 2020, p. 5).

- **Supporting a civic deliberative community:** Schuler (2020, p. 8) writes that "citizens need civic culture and they need civic infrastructure." Such an approach could be used to create platforms for citizen exchange, deliberation and voting on local developments, petitions, suggestions, and similar ways to help citizens exert power through self-organization. Imagine apps that integrate access to administrative services and have communication functions that enable citizens to ask for help, discuss with each other, propose changes, and provide feedback.

The above list may appear mundane, and it is only a brief sketch that can surely be refined through participatory development frameworks. What is essential to take this further, however, is to approach the division of e-government services through a lens focused on *digitally aided sovereignty*.

5 CONCLUSION

This contribution is an attempt at elaborating on the utility of adopting the notion of digitally aided sovereignty as a guide for the e-government transformation. We have tried to elucidate that a novel stance towards the digitalization of public administration may be necessary to aid in realizing practical and increased participation and consultation. Considering the current state of e-democracy and e-government, this stance needs to be pragmatic instead of overly positivistic. We have sketched examples of practices that use digital technology to provide public services and take citizens seriously in their democratic capacity—examples of what we call digitally aided sovereignty in e-government. While keeping the goal of reducing social inequity firmly in sight, ironically, this may mean more rather than fewer encounters with public administration. In fact, this is where we see untapped potential for increased participation.

6 REFERENCES

1. Chadwick, A., & May, C. (2003). Interaction between States and Citizens in the Age of the Internet: "e-Government" in the United States, Britain, and the European Union. *Governance*, *16*(2), 271–300. https://doi.org/10.1111/1468-0491.00216

2. Christiansen, J.-A. (2010). E-Democracy. In H. Rahman (Ed.), *Handbook of Research on E-Government Readiness for Information and Service Exchange* (pp. 271–294). IGI Global. https://doi.org/10.4018/978-1-60566-671-6.ch014

3. Couture, S., & Toupin, S. (2019). What does the notion of "sovereignty" mean when referring to the digital? *New Media & Society*, *21*(10), 2305–2322. https://doi.org/10.1177/1461444819865984

4. Digitalisierung gestalten – Umsetzungsstrategie der Bundesregierung (p. 224). (2020).

5. European Commission. (2015). *A Digital Single Market Strategy for Europe* (Communication, p. 192). European Commission.

6. European Parliament. Directorate General for Parliamentary Research Services. (2017a). *Prospects for e-democracy in Europe: Literature review, Part I.* Publications Office. https://data.europa.eu/doi/10.2861/49654

7. European Parliament. Directorate General for Parliamentary Research Services. (2017b). *Prospects for e-democracy in Europe: Study summary.* Publications Office. https://data.europa.eu/doi/10.2861/201697

8. Floridi, L. (2015). Tolerant Paternalism: Pro-ethical Design as a Resolution of the Dilemma of Toleration. *Science and Engineering Ethics*. https://doi.org/10.1007/s11948-015-9733-2

9. Floridi, L. (2020). The Fight for Digital Sovereignty: What It Is, and Why It Matters, Especially for the EU. *Philosophy & Technology*, *33*(3), 369–378. https://doi.org/10.1007/s13347-020-00423-6

10. Friesike, S., & Sprondel, J. (2022). Träge Transformation. Welche Denkfehler den digitalen Wandel blockieren. Reclam.

11. Germany's Federal Ministry of Education and Research. (2019, January 3). *Richtlinie zur Förderung von Forschung und Entwicklung auf dem Gebiet "Mensch-Technik-Interaktion für digitale Souveränität"*. https://www.bmbf.de/bmbf/shareddocs/bekanntmachungen/de/2019/03/2355_bekanntmachung

12. Gesellschaft für Informatik (GI) & Verband Deutscher Elektrotechniker (VDE). (2000). Electronic Government als Schlüssel zur Modernisierung von Staat und Verwaltung. Ein Memorandum des Fachausschusses.

13. Grönlund, Å. (2010). Ten Years of E-Government: The 'End of History' and New Beginning. In M. A. Wimmer, J.-L. Chappelet, M. Janssen, & H. J. Scholl (Eds.), *Electronic Government* (Vol. 6228, pp. 13–24). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-14799-9_2

14. Madsen, C. Ø., Berger, J. B., & Phythian, M. (2014). The Development in Leading e-Government Articles 2001-2010: Definitions, Perspectives, Scope, Research Philosophies, Methods and Recommendations: An Update of Heeks and Bailur. In M. Janssen, H. J. Scholl, M. A. Wimmer, & F. Bannister (Eds.), *Electronic Government* (Vol. 8653, pp. 17–34). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-44426-9_2

15. *Mensch-Technik-Interaktion für digitale Souveränität*. (2020). Mensch-Technik-Interaktion Für Digitale Souveränität – Geförderte Projekte. https://www.interaktive-technologien.de/foerderung/bekanntmachungen/digisou

16. Mergel, I. (2021). Digital Transformation of the German State. In S. Kuhlmann, I. Proeller, D. Schimanke, & J. Ziekow (Eds.), *Public Administration in Germany* (pp. 331–355). Springer International Publishing. https://doi.org/10.1007/978-3-030-53697-8_19

17. Merkel, W. (2020, January 4). Who is the sovereign? *Berlin Social Science Center, Coronavirus and Its Impact*. https://www.wzb.eu/en/research/corona-und-die-folgen/wer-ist-der-souveran

18. Moerel, E. M. L., & Timmers, P. (2021). *Reflections on Digital Sovereignty* (EU Cyber Direct, Research in Focus Series). https://ssrn.com/abstract=3772777

19. Mokyr, J. (2000). Innovation and Its Enemies: The Economic and Political Roots of Technological Inertia. In M. Olson & S. Kähköhnen (Eds.), *A Not-so-dismal Science*. Oxford University Press. https://doi.org/10.1093/0198294905.001.0001

20. Pernice, I. (2016). E-Government and E-Democracy: Overcoming Legitimacy Deficits in a Digital Europe. 2016–01, 29. http://dx.doi.org/10.2139/ssrn.2723231

21. Pohle, J., & Thiel, T. (2020). Digital sovereignty. *Internet Policy Review*, *9*(4). https://doi.org/10.14763/2020.4.1532

22. Pohle, J., & Thiel, T. (2021). Digitale Souveränität: Von der Karriere eines einenden und doch problematischen Konzepts. In C. Piallat (Ed.), *Digitale Gesellschaft* (1st ed., Vol. 36, pp. 319–340). transcript Verlag. https://doi.org/10.14361/9783839456590-014

23. Pratchett, L. (2006). *Understanding e-democracy developments in Europe* (CAHDE(2006) 2 E; Project "Good Governance in the Information Society"). Council of Europe, Directorate General of Political Affairs.

24. Schmid, A. M., Recker, J., & vom Brocke, J. (2017). *The Socio-Technical Dimension of Inertia in Digital Transformations*. Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2017.583

25. Schuler, D. (2020). Can Technology Support Democracy? *Digital Government: Research and Practice*, *1*(1), 1–14. https://doi.org/10.1145/3352462

26. Stowers, G. (2013). *The Use of Data Visualization in Government* (p. 49). IBM Center for The Business of Government.

27. Winkel, O. (2007). Electronic Government in Germany – a key future prospect, but expectations are exaggerated. In K. Zapotoczky & C. Pracher (Eds.), *Administration Innovative* (pp. 163–186).

Proceedings of the Weizenbaum Conference 2022:
Practicing Sovereignty. Interventions for Open Digital Futures

MIGRANTS' IMAGINARIES AND AWARENESS OF DISCRIMINATION BY ARTIFICIAL INTELLIGENCE

A CONCEPTUAL FRAMEWORK FOR ANALYSING DIGITAL LITERACY

Sūna, Laura
Institute for Media Studies
University of Siegen
Siegen, Germany
laura.suna@uni-siegen.de

KEYWORDS

digital literacy; imaginaries about artificial intelligence; folk theories.

DOI: 10.34669/wi.cp/4.2

ABSTRACT

This paper asks what skills migrants need to be able to deal with artificial intelligence (AI) technologies in a self-determined way in their everyday lives. We propose a conceptual framework to empirically identify migrant's awareness and perceptions of possible discrimination through AI. Following Bucher (2017, 40), we argue that by experiencing AI systems in their digital environments, people develop AI imaginaries that shape their attitudes, interactions, and practices with AI. We assume that experiences of discrimination evoke affects, feelings, and emotions that at first glance are not associated with AI technologies. The paper provides relevant research questions that address AI imaginaries. In addition to studying knowledge about and perceptions of AI, research should increasingly focus on users' attitudes towards AI, their evaluations of AI, and their feelings, emotions, and affects related to AI. Subsequently, we elaborate on dimensions of digital literacy based on these AI imaginaries. Finally, we will describe the digital skills that are necessary to confidently cope with discrimination by AI technologies.

1 INTRODUCTION: MIGRANTS' EXPERIENCES OF DISCRIMINATION BY AI TECHNOLOGIES

From the research literature on self-learning systems, we know that various artificial-intelligence-based technologies reproduce bias and thus reinforce inequality and exclusion (Tulodziecki, 2020). For example, Lopez (2021) describes how certain groups are rendered invisible, overly visible, or distorted due to a "racial bias." At the same time, social bias and stereotypes are reproduced as structural inequalities within the society, while they are emphasized and reinforced by algorithms and artificial intelligence (AI). Nevertheless, studies have shown that users often do not perceive the algorithmic curation of various AI-driven services or do not see it as problematic or discriminatory (MeMo:KI, 2020; Swart 2021). For a self-determined life with digital media, however, an informed, reflective, critical approach to AI is necessary (Digitales Deutschland, 2021). This is especially true for groups that are particularly affected by inequality, such as migrants. Confident use of AI technologies is also crucial for migrants' participation practices.

We therefore ask why the discrimination caused by AI is scarcely perceived by the population. One of the reasons may be the invisibility and opaqueness of AI technologies for laypeople and semi-professional users of digital systems. Invisibility also makes empirical research on possible experiences of discrimination difficult. At the same time, discrimination can be latent; those who are discriminated against may perceive it unconsciously. These aspects also make it difficult to research experiences of discrimination in the digital world. So how can we grasp an encounter with AI technology when it is perceived unconsciously and the technology is not always visible? One possible approach is to explore subjective everyday life theories about the use of a given technology: in our case AI. However, even if most users do not understand the technical aspects of algorithms and AI, they can, nevertheless, or even for that very reason, form an understanding of it. In the research literature, we find several terms for these everyday life theories: for example, folk theories, lay theories, mental models or more concretely "users' intuitive theories about the composition of their Facebook News Feeds," or "algorithmic expertise and algorithmic gossip" (Rader & Gray, 2015; Bishop, 2019; Dogruel et al., 2020). Such subjective everyday life theories do not necessarily correspond to technological developments; they are also constantly updated and if necessary refuted by the user. They describe the imaginaries about AI that people develop.

Studies in the field of computer-human interaction focus their analyses on algorithms and on how aware the general digital-media-using population is of them. We are interested in a wider field of user perception and interaction with AI technologies. Algorithms are an essential part of many AI systems, but AI encompasses many other technologies as well. Algorithmic decision-making in online applications is certainly one of the most familiar AI technologies. In everyday language, the

term algorithm is often used synonymously with the term artificial intelligence, as different studies about the perception of AI show (Digitales Deutschland, 2021b). However, we are less interested in the technological understanding and definitions of AI; we will instead focus on migrants' attributions of meaning, evaluations, and practices of using AI technologies.

In the following, we will present a conceptual approach to empirically study migrants' possible experiences of discrimination by AI technologies. We will argue that analyzing subjective everyday life theories can render this study possible. These subjective theories will help us analyze migrants' perceptions and their coping mechanisms for dealing with discrimination by AI systems. In addition, the AI imaginaries make it possible to determine the dimensions of digital literacy that migrants have and need to deal with AI technologies in a self-determined way in everyday life. Various coping mechanisms regarding discrimination by AI can then be linked to these.

2 SUBJECTIVE EVERYDAY LIFE THEORIES ABOUT ARTIFICIAL INTELLIGENCE

What exactly are subjective everyday life theories and what do they reveal about the role of AI in migrants' lives and experiences of discrimination? People develop folk theories to explain aspects of life that are complex in their structure—be they technical or other features. Such mental models represent *a subjective idea of what* AI *is*, what it *can do*, and *what it should do*. These theories serve as a symbolic resource for practices of using AI technologies (Ytre-Arne & Moe, 2021, p. 810). Subjective lay theories are intuitive ways of thinking about the function and structure of things, and especially about technologies. They shape the practices of using and experiencing them (Siles et al., 2020, p. 2) and help people to interact with the technology. Subjective lay theories also contain strategies of action and thus include general solutions for dealing with technology, which can be valid and exemplary in various everyday situations. Ytre-Arne and Moe (2021) highlight the "value of folk theories not just in guiding behavior, but also in making sense of experiences, generating inference and steering learning about the world." (Ytre-Arne & Moe, 2021, p. 811) Furthermore, the respondents' statements about their folk theories can be contradictory. This is not surprising, as they are often not as well-formed or thoroughly tested as scientific theories. Moreover, people are constantly developing new perspectives on technologies through continuous interaction with them (Rader & Gray, 2015, p. 177).

We thus argue that subjective everyday life theories build on people's *perceptions* and *knowledge* about, *attitudes* towards, and *emotional* and *affective* evaluations of AI. These theories emerge in a relational process: They are developed, adapted, and changed through the experience of using AI systems. An unaddressed issue in previous research is that the affective dimension is rarely

considered when analyzing the imagining and use of AI systems (with the exception of the studies by Bucher, 2017; 2018 and Swart 2021). We argue that it is, however, an important facet of attitudes towards AI and its use.

We propose to analyze people's personal AI stories to explore what subjective everyday life theories migrants develop about AI and what these theories uncover about digital literacy and discriminatory experiences with AI. These are "stories about situations and disparate scenes that draw algorithms and people together." (Bucher, 2017, p. 30) The focus is on what people imagine and associate with AI and algorithms. Bucher (2017) calls it the "algorithmic imaginary." Such imaginaries of AI are strongly shaped by feelings and emotions. What people experience when using AI systems is not a mathematical equation but rather the moods, affects and feelings evoked by AI systems. We are interested in "beliefs people form about algorithms not as right or wrong but as a point of access for understanding how and when algorithms matter." (Bucher, 2018, pp. 97–8) This helps us to understand the social power of AI. Related to this are questions regarding what emotions and feelings are evoked by the experience of inequality and discrimination by algorithmic systems. Similarly, Ytre-Arne and Moe (2021, p. 810) argue that the subjective construction of the meaning of AI and algorithms is an interpretive process, it is less about technical knowledge than about the negotiation and attribution of meanings and evaluations. AI imaginaries can emerge in a collaborative way, and they are negotiated in the context of cultural and societal beliefs. Moreover, they are imprecise and can embody cognitive biases (Bishop, 2019, 2593; Ngo & Krämer, 2021, p. 3). Imaginaries can be seen as the "outcome of individual and collective sensemaking activities resulting in shared ideas about technology, including fears, hopes, and expectations." (Kazansky & Milan, 2021, p. 364) Furthermore, subjective everyday life theories and AI imaginaries rely on different discourses about AI. Mager and Katzenbach (2021, p. 223) emphasize that AI imaginaries are increasingly dominated by discourses developed by technology companies, followed by discourses from science fiction and the media.

Quantitative studies (Fischer & Petersen, 2018) have shown that the population's overall knowledge and awareness of AI is relatively low. Several studies have demonstrated that knowledge about and awareness of AI can be better explored through qualitative research. For example, in an open-ended interview, the interviewer can connect with the interviewee's concrete experiences of use and interaction with AI systems (Dogruel et al., 2020, p. 13). Hargittai and colleagues (2020, p. 767) asked users about their "individual perception and understanding of online processes with algorithms" instead of asking directly about their knowledge about algorithms. In this context, we argue that whether people have explicit knowledge about the specific functioning of the systems is of secondary importance for research into possible discrimination by AI systems. Rather, we are interested in how

users perceive these systems, how they feel them, interact with them, evaluate them, and how they adapt their practices in this context. How self-determined and sovereign do they perceive their interaction with AI systems to be, or to what extent can they shape this interaction in a self-determined way.

3 ADDRESSING AI IMAGINARIES AND DIGITAL LITERACY

In the following section, we will identify useful research questions for analyzing self-determined interactions with AI technologies. As previously described, it can be helpful to use qualitative methods to examine AI imaginaries and possible discriminatory experiences with AI. These methods allow us to "capture more intuitive, tacit forms of knowledge and make it possible to explore openly how people understand and engage with algorithms" (Swart, 2021, p. 3). In open-ended interviews or group discussions, researchers can uncover the interviewees' indirect experiences with AI systems, which they have often not reflected upon. It is helpful to connect to the experience of using and interacting with AI systems in everyday life (Hargittai et al., 2020, p. 771; Dogruel et al., 2020, p. 13). Siles et al. (2020, p. 4) have argued that group discussions are "ideal for exploring the social nature of folk theories, that is, how they form as people share them with others." When group discussions are combined with the method of rich pictures (Bell & Morse, 2013), which involve drawings about the role of the technology in everyday life made by individuals, researchers can understand the unstated and self-evident nature of users' knowledge of AI systems. Two more useful methods are the think-aloud method (Charters, 2003) and the walkthrough method (Light et al., 2018), which means that an individual goes through their device together with the researcher and describes typical practices of the use of AI technologies. These drawings and stories can be used as starting points for group discussions or personal interviews (Siles et al., 2020). In their studies, Bucher (2017, p. 38) and Dogruel et al. (2020, p. 4) have shown that the power of algorithms and AI becomes particularly visible when users face problems and irritations due to algorithms or AI. This often negatively associated interaction with algorithms makes the penetration of their social environment by AI systems visible. Conversations about these experiences have provided promising results.

4 RESEARCH QUESTIONS ABOUT AI IMAGINARIES

We propose the following steps for empirically identifying AI imaginaries. It is useful to start with overall questions about the respondent's general experience of using AI systems. When it comes to identifying important aspects of everyday life theories of AI, respondents' ideas about AI and attitudes towards it, their evaluations of AI, and their feelings, emotions, and affects related to AI are important—in addition, of course, to their awareness of, knowledge about, and perception of AI. In

this context, the following research questions have emerged (several of them have been analyzed in: Bucher 2017; Pedersen, 2019; Dogruel et al., 2020; Siles et al., 2020; Hargittai et al., 2020; Swart 2021; Ytre-Arne & Moe, 2021):

1. *Awareness, opinion, perception, and attitudes about AI technologies*: How do users make sense of the mostly opaque AI systems? To what extent are migrants aware of AI in their media use? How are these AI technologies integrated into people's everyday lives? How do users think that algorithmic recommendations work, for example? How do users perceive the role of AI technologies in their lives? What positive and negative aspects of algorithms do users perceive? To what extent do they find them discriminatory?
2. *Emotions and feelings towards AI*: If users do recognize the existence and role of AI, how do they feel about it? What emotions and feelings might refer to a discriminatory experience?
3. *Evaluation of and reflection on AI*: How much do users reflect on the impact of AI systems when using the internet and digital media? Do they understand when and how AI may influence their actions? How do users adapt their everyday practices in response to critical evaluation? Do they reflect on changes in the technology over time and how it changes their use of AI systems? To what extent do respondents reflect on ethical aspects of their individual use of AI systems.
4. *Media practices adapted to AI imaginaries*: How do users adapt their media practices with AI systems within the context of their imaginaries, awareness, knowledge, evaluation, and emotions? How self-determined are these practices?

What can these aspects of AI imaginaries reveal about sovereign living in general and about practices of coping with possible discrimination by AI?

5 ADDRESSING DIGITAL SOVEREIGNTY AND DIGITAL LITERACY

Furthermore, we relate the skills described in various media and digital literacy models (Hobbs, 2010; Livingstone et al., 2005) to AI imaginaries and thus describe the competence requirements that the subjects must face if they want to participate sovereignly in a democratic society shaped by digitalization. Based on Müller et al. (2020, p. 32), we define digital sovereignty as the skills and opportunities of a person to shape their own life in a competent, self-determined, and secure way when using or depending on digital media. Digital sovereignty is relational and is shaped not only by individual conditions but also by technical, legal, and social ones.

The Digital Germany research network (Digitales Deutschland, 2021) highlights the following skills that can enable a sovereign way of living in a digitalized society: *instrumental skills* (the skills necessary to use digital media), *cognitive and critical-reflexive skills* (knowledge about AI systems

and how to evaluate them), *creative skills* (concerning the self-determined, independent (re)design of digital media and systems), and *affective and social skills* (being able to react emotionally and socially appropriately to media content and AI systems). The affective and social dimension is neglected in many models of AI or algorithm literacy (cf. the comprehensive systematizations by Dogruel, 2021; Long & Magerko, 2020). If we relate the digital literacy dimensions to the research questions we have formulated, the following connection emerges (see Table 1): The first set of questions analyses the cognitive and critical-reflective aspects, such as awareness, opinion, perception, knowledge, and attitudes towards AI. The second question highlights the affective and social dimension of dealing with AI and possible discrimination. The third set of questions describes the critical evaluation of AI technologies. The last point focuses on creative practices and coping strategies in dealing with AI.

Question	Digital literacy dimensions
Awareness, opinion, perception, attitudes about AI technologies	Cognitive and critical-reflexive skills
Emotions and feelings toward AI	Affective and social skills
Evaluation of and reflection on AI	Critical evaluation skills
Media practices adapted to AI imaginaries	Creative coping strategies with the impact of AI—creative skills

Table 1. Digital literacy and AI imaginaries.

In the context of digital sovereignty, we particularly want to highlight the productive power of AI imaginaries. Even though users' agency around AI systems is substantially limited by platform structures, they develop different coping strategies (Swart, 2021). Nevertheless, subjective everyday life theories describe people's productive interaction with AI (Bucher 2017, p. 41; Bishop 2019, p. 2592). Based on their affectively informed experience with the algorithm, people try to change it, adapt to it, and make it useful for themselves. Ytre-Arne and Moe (2021, p. 809) also emphasize that unexpected experiences with AI promote user's activity and performativity. This can be particularly interesting for research on self-efficacy in the face of discrimination by AI systems. Everyday user encounters with AI can contribute to the development of resistance practices (Velkova & Kaun, 2021, p. 525).

6 CONCLUSION

Our starting point was the question of what skills and abilities migrants need to be able to deal sovereignly with AI systems in their everyday life. Linked to this was the question of how to best explore imaginaries and experiences of AI and how to analyze potential experiences of discrimination. Our proposed approach was to explore AI imaginaries as subjective everyday life theories to describe people's perceptions and knowledge about, attitudes towards, and emotional and affective evaluations of AI. Furthermore, the AI imaginaries describe what digital literacy the migrants have as well as what is necessary for them to lead a sovereign life in the digitalized world. Digital literacy refers to possible skills and abilities to deal with discrimination experiences with AI: in the sense of being able to recognize and counter discrimination. We argue that especially the affective, social, critically reflexive, and creative skills and abilities are significant for a sovereign way of living.

7 ACKNOWLEDGMENTS

This publication is part of the project “Digital Germany—Monitoring the population's digital competence,” funded by the German Federal Ministry for Family Affairs, Senior Citizens, Women and Youth.

REFERENCES

1. Bell, Simon, und Stephen Morse. 2013. “How People Use Rich Pictures to Help Them Think and Act”. *Systemic Practice and Action Research* 26 (4): 331–48. https://doi.org/10.1007/s11213-012-9236-x.
2. Bishop, Sophie. 2019. “Managing Visibility on YouTube through Algorithmic Gossip”. *New Media & Society* 21 (11–12): 2589–2606. https://doi.org/10.1177/1461444819854731.
3. Bucher, Taina. 2017. “The Algorithmic Imaginary: Exploring the Ordinary Effects of Facebook Algorithms”. *Information, Communication & Society* 20 (1): 30–44. https://doi.org/10.1080/1369118X.2016.1154086.
4. ———. 2018. *If...then: algorithmic power and politics*. New York: Oxford University Press.
5. Charters, Elisabeth. 2003. “The Use of Think-aloud Methods in Qualitative Research. An Introduction to Think-aloud Methods”. *Brock Education* 12 (2): 68–82.
6. Cotter, Kelley, und Bianca C. Reisdorf. 2020. “Algorithmic Knowledge Gaps: A New Horizon of (Digital) Inequality”. *International Journal of Communication* 14 (0): 21. https://ijoc.org/index.php/ijoc/article/view/12450.
7. Digitales Deutschland. 2021. Rahmenkonzept Digital- und Medienkompetenz, accessed February 15, 2022. https://digid.jff.de/rahmenkonzept/.
8. Digitales Deutschland. 2021b. Data Dashboard. Accessed August 9, 2022. https://digid.jff.de/data-dashboard/.
9. Dogruel, Leyla, Dominique Facciorusso, and Birgit Stark. 2020. “I’m Still the Master of the Machine. Internet Users’ Awareness of Algorithmic Decision-Making and Their Perception of Its Effect on Their Autonomy”. *Information, Communication & Society*, 1–22. https://doi.org/10.1080/1369118X.2020.1863999.
10. Dogruel, Leyla. 2021. “What Is Algorithm Literacy? A Conceptualization and Challenges Regarding Its Empirical Measurement“. In *Algorithms and Communication*, edited by Monika Taddicken and Christina Schumann, 9:67–93. Digital Communication Research. Berlin. https://doi.org/10.48541/dcr.v9.3.
11. Fischer, Sarah, and Petersen, Thomas. 2018. “Was Deutschland über Algorithmen weiß und denkt. Ergebnisse einer repräsentativen Bevölkerungsumfrage“. Bertelsmann Stiftung. Accessed February 15, 2022. https://doi.org/10.11586/2018022.
12. Hargittai, Eszter, Jonathan Gruber, Teodora Djukaric, Jaelle Fuchs, and Lisa Brombach. 2020. “Black Box Measures? How to Study People’s Algorithm Skills”. *Information, Communication & Society* 23 (5): 764–75. https://doi.org/10.1080/1369118X.2020.1713846.
13. Hobbs, Renee. 2010. Digital and media literacy: A plan of action. Washington, DC: Knight Foundation and Aspen Institute. Accessed February 15, 2022. https://knightfoundation.org/wp-content/uploads/2019/06/Digital_and_Media_ Literacy_A_Plan_of_Action.pdf.
14. Kazansky, Becky, and Stefania Milan. 2021. “‘Bodies Not Templates’: Contesting Dominant Algorithmic Imaginaries.” *New Media & Society* 23, No. 2: 363–81. https://doi.org/10.1177/1461444820929316.
15. Light, Ben, Burgess, Jean, & Duguay, Stefanie. 2018. “The walkthrough method: An approach to the study of apps.” *New Media & Society*, 20(3), 881–900. https://doi.org/10.1177/1461444816675438.

16. Livingstone, Sonia, Elizabeth van Couvering, and Nancy Thumim. 2005. *Adult Media Literacy: A Review of the Research Literature*. London: Ofcom.

17. Long, Duri, and Brian Magerko. 2020. "What Is AI Literacy? Competencies and Design Considerations". In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16. Honolulu HI USA: ACM. https://doi.org/10.1145/3313831.3376727.

18. Lopez, Paola. 2021. "Diskriminierung durch Data Bias. Künstliche Intelligenz kann soziale Ungleichheiten verstärken." *WZB Mitteilungen.* Heft 171: 26-28.

19. Mager, Astrid, and Christian Katzenbach. 2021. "Future Imaginaries in the Making and Governing of Digital Technology: Multiple, Contested, Commodified." *New Media & Society* 23, No. 2: 223–36. https://doi.org/10.1177/1461444820929321.

20. MeMo:KI. 2020. Meinungsmonitor Künstliche Intelligenz. Künstliche Intelligenz und Diskriminierung. Factsheet Nr. 2 (August). Accessed February 15, 2022. http://www.cais.nrw/wp-94fa4-content/uploads/2020/08/Factsheet-2-KI-und-Diskriminierung.pdf.

21. Müller, Jane, Mareike Thumel, Katrin Potzel, and Rudolf Kammerl. 2020. "Digital Sovereignty of Adolescents". *MedienJournal* 44 (1): 30–40. https://doi.org/10.24989/medienjournal.v44i1.1926.

22. Ngo, Thai, and Krämer, Nicole. 2021. "It's just a recipe? – Comparing expert and lay user understanding of algorithmic systems." *Technology, Mind, and Behavior*, *2*(4): 1–10. https://doi.org/10.1037/tmb0000045.

23. Pedersen, Emily. 2019. "'My Videos are at the Mercy of the YouTube Algorithm': How Content Creators Craft Algorithmic Personas and Perceive the Algorithm that Dictates their Work". Technical Report, University of California at Berkeley.

24. Rader, Emilee, and Rebecca Gray. 2015. "Understanding User Beliefs About Algorithmic Curation in the Facebook News Feed". In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 173–82. CHI '15. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2702123.2702174.

25. Siles, Ignacio, Andrés Segura-Castillo, Ricardo Solís, and Mónica Sancho. 2020. "Folk Theories of Algorithmic Recommendations on Spotify: Enacting Data Assemblages in the Global South". *Big Data & Society.* January–June: 1–15. https://doi.org/10.1177/2053951720923377.

26. Swart, Joelle. 2021. "Experiencing Algorithms: How Young People Understand, Feel About, and Engage With Algorithmic News Selection on Social Media". *Social Media + Society*, *7*(2), 1–11. https://doi.org/10.1177/20563051211008828.

27. Tulodziecki, Gerhard. 2020. Künstliche Intelligenz und Medienpädagogik. Zwischen Utopie und Dystopie. In *Medienpädagogische Perspektiven für die digitale Gesellschaft,* edited by Angelika Beranek, Sebastian Ring, Martina Schuegraf. 1-15. München: koPaed.

28. Velkova, Julia, and Anne Kaun. 2021. "Algorithmic Resistance: Media Practices and the Politics of Repair". *Information, Communication & Society* 24 (4): 523–40. https://doi.org/10.1080/1369118X.2019.1657162.

29. Ytre-Arne, Brita, and Hallvard Moe. 2021. "Folk Theories of Algorithms: Understanding Digital Irritation". *Media, Culture & Society* 43 (5): 807–24. https://doi.org/10.1177/0163443720972314.

Proceedings of the Weizenbaum Conference 2022:
Practicing Sovereignty. Interventions for Open Digital Futures

EUROPE'S DIGITAL SOVEREIGNTY

AN INTERNATIONAL POLITICAL ECONOMY CONCEPTUAL APPROACH

Kreutzer, Stephan
Technopolis Group
Berlin, Germany
stephan.kreutzer@technopolis-group.com

Vogelsang, Manuel Molina
Fraunhofer IMW
Leipzig, Germany
manuel.molina.vogelsang@imw.fraunhofer.de

KEYWORDS

digital sovereignty; economic policy; international political economy, market competition; technology companies; data economy

DOI: 10.34669/wi.cp/4.3

ABSTRACT

This paper looks at conceptual approaches to digital sovereignty from an international political economy perspective, focusing on the state level. We consider the implications of the rise of the data economy and analyze different economic policy approaches to restoring and preserving Europe's digital sovereignty from market liberal and industrial policy perspectives. We conclude that networked sovereignty can optimally be attained by supporting the emergence and success of homegrown technology companies in a globalized data economy. Digital sovereignty can best be achieved by policy makers using a mix of market liberal and more proactive industrial policy instruments. The liberal focus on framework conditions is useful in refocusing policy makers' efforts on deepening the EU single market, while the industrial policy approach can be a suitable way of funding pilot projects in early-stage technology areas in partnership with industry and setting rules for newly emerging markets. State action is also necessary to avoid monopolies.

1 INTRODUCTION

One of the most salient phenomena of today is the digital transformation of society. The impact these developments are having on the competitiveness of businesses, market structures, and global value networks are of geopolitical relevance (Brynjolfsson and Saunders 2010; Rumana Bukht and Richard Heeks 2017; van de Velde et al. 2015). From a European Union (EU) perspective, the rise of data-driven business models introduced by digital platform companies[3] in the United States and, to a lesser extent, in China, may conflict with the political goals of protecting citizens' privacy and enhancing the competitiveness of homegrown companies. Due to the importance of digital technologies—such as semiconductors, cloud computing, and artificial intelligence—for various industries, such conflicts have implications for the technological sovereignty of the EU (Braud et al. 2021; Bauer and Erixon 2020; Bendiek and Neyer 2020; Edler et al. 2020).

The aim of this paper is to compare different approaches rooted in economic theory to identify a practical approach to answering the following question: How can Europe restore and preserve its digital sovereignty? Conceptually, the paper contributes to the academic debate on the geopolitical and international political economy perspective on digital sovereignty: How can digital sovereignty be defined at the level of states and supranational entities (such as the EU)? What challenges and potential opportunities does the rise of digital platform companies present for political actors? What (combination of) economic policy instruments supporting digital sovereignty show most promise? In tackling these questions, the paper touches upon regulatory and technological aspects and questions of autonomy, control, and authority in a globalized data economy.

In Section 2, we summarize the academic debate on "digital sovereignty" and define the term at the state level. In Section 3, we outline the characteristics of the data economy and the European position vis-à-vis major competitors. We analyze economic policy approaches to restoring and preserving digital sovereignty at the EU level in Section 4. Finally, we conclude by summarizing the results and outlining further research questions.

2 CONCEPTUAL CONSIDERATIONS

The concept of "sovereignty" has varying definitions and has been used in different contexts throughout history. Today, the term digital sovereignty is widely used in the political debate (Krasner 2012; Korff 1923; Biersteker 2002, 1999; Couture and Toupin 2019; Hummel et al. 2021; Pohle and Thiel 2020). To analyze the interlinkages between the geopolitical and economic dimensions of

[3] In the following, we understand technology and digital platform companies as encompassing those enterprises whose business model is highly dependent on R&D and the use of digital technologies.

digital sovereignty at the state level, we suggest adopting an international political economy viewpoint in the following discussion. Therefore, we conducted a semantic search for "digital sovereignty" in the titles, abstracts, and keywords recorded in the bibliographic databases Scopus, Web of Science, and Dimensions.ai.

Figure 1 shows that the debate on "digital sovereignty" is a relatively new one. The first documented publications date back to 2013; since 2018, the volume of publications has increased substantially. The analysis shows that fewer publications were listed in the academic databases Scopus and Web of Science than in Dimensions.ai, which also covers reports, working papers, and policy documents. Nevertheless, all three bibliographic databases show the same dynamics.

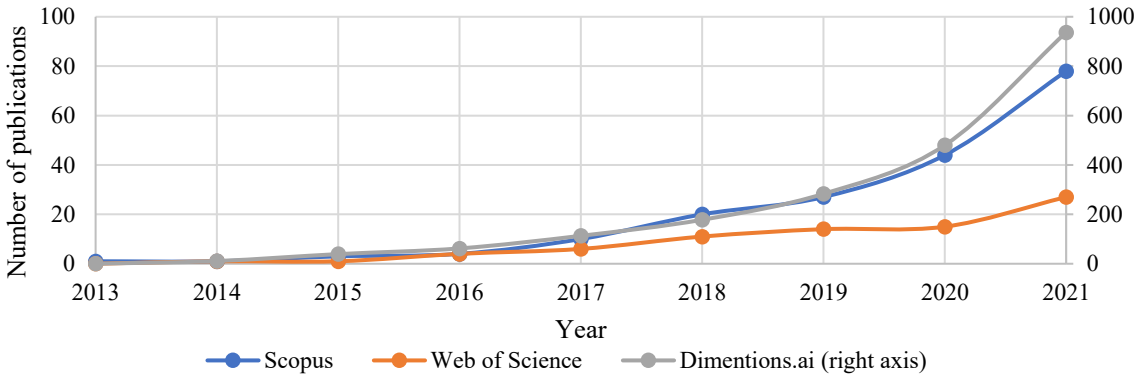


Figure 1. Development of publications for the keyword "digital sovereignty" in different bibliographic databases; own depiction based on Scopus, Web of Science, and Dimentions.ai

Furthermore, most of the publications are in computer sciences and social sciences (see Table 1). Most publications are from European (Germany, France, Finland, UK) and to a lesser extent US authors.

Top subject area	Scopus	Web of Science	Dimensions.ai
1st	Computer Science (39)	Computer Science Information Systems (6)	Studies in Human Society (307)
2nd	Social Sciences (38)	Computer Science Theory Methods (6)	Information and Computing Sciences (185)
3rd	Engineering (14)	Communication (5)	Law and Legal Studies (144)
4th	Arts and Humanities (11)	Engineering Electrical Electronic (4)	Economics (62)
5th	Business, Management and Accounting (8)	Telecommunications (3)	Commerce, Management, Tourism and Services (52)

Figure 2. Most relevant subject areas of publication on "digital sovereignty"; number of publications between 2013 and 2021 in parentheses; own elaboration based on Scopus, Web of Science, and Dimentions.ai

Regardless of the scientific or geographical background, the concept of "digital sovereignty" refers to an individual's capacity to pursue their own goals in a self-determined way and without being limited by access to key digital technologies and competences (Pohle and Thiel 2020; Lambach 2020; Wittpahl 2017; VDE 2020). It is widely acknowledged that digital sovereignty can be used at the individual, organizational, national, and even supranational levels.

At the **individual level**, the term usually refers to citizens' right to privacy and data protection but also to their ability and competences to use digital technologies in a self-determined way. This is relevant for the geopolitical dimension of sovereignty insofar as broad acceptance of digital technologies is needed for these to successfully be taken up in society. Societal take-up, in turn, facilitates the creation of a competitive technology base. Moreover, most data-driven business models rely on citizens' willingness to share their personal data.

At the **organizational level**, digital sovereignty can be assessed from the point of view of companies and other organizations. Here, aspects such as cybersecurity or intellectual property and the control of data are of major concern. Questions of competitiveness and value chains are equally relevant in this context.

At the **state level**, in focus here, the term touches upon regulation. Here, data sovereignty can translate into technological leadership and into economic competitiveness and geopolitical power. As digitalization reaches ever more industries and knowledge areas, the question of who has access to and controls the data impacts on the competitiveness of entire economies. Global leadership in key technology areas and innovation does not only mean locating high-added-value activities and

competitive industries at home but also results in geopolitical power abroad and makes it possible to shape global economic and social governance and rules.

There is also the question of what *constitutes* sovereignty and how it can best be achieved in the globalized data economy. Digital sovereignty is widely considered to solve a set of interlinked problems in various policy areas (Bendiek and Neyer 2020; Pohle and Thiel 2020; VDE 2020). At the state level, it is argued that neither autonomy, nor autarky, nor heteronomy are viable options for achieving and preserving digital sovereignty. Digital sovereignty is best achieved through networking and diversification, and global interconnectedness and interdependence (Syuntyurenko 2015). We conclude that digital sovereignty manifests itself in the interaction between economic, technological, and political actors.

The bibliographic analysis shows that the debate focuses more on the individual and organizational level than the state level. We find little evidence of conceptual or empirical studies on the best choice of economic policy approaches to preserve digital sovereignty at the state level. To help fill this gap, we will first look at the features of the data-driven economy and then summarize the key notions of each of these economic policy approaches. This research focus implies that we regard digital sovereignty as a normative term, that is, as a status that is considered desirable.

3 APPLICATION: THE CASE OF EUROPEAN DIGITAL SOVEREIGNTY

The EU as a supranational entity faces challenges to its digital sovereignty (Floridi 2020). Until recently, Europe was able to shape and influence global rules for the data economy, but it was also able to influence rulemaking in other technological and economic areas. By making access to its market of more than 450 million people contingent on compliance with its own rules, it has often led technology companies to adopt global policies that are in line with EU regulation. This has been termed the “Brussels effect” (Bradford 2020). The most prominent example is the General Data Protection Regulation (GDPR) which has been mirrored in other jurisdictions, most notably in California (Sirota, 2019; Voss and Houser 2019; Baik 2020). However, as Europe’s share of global GDP shrinks, its power to influence global rules will do so as well in the absence of homegrown digital platform companies and in the wake of rising competition from the United States, China, and elsewhere.

This gives rise to the question of why Europe has so far failed to create digital technology companies that could shape the global data economy. The data economy is characterized more than other industries by scale and network effects (Rochet and Tirole 2003; Katz 1994; Alt and Zimmermann 2019). In digital markets, there is a tendency towards oligopolies; indeed, in more narrowly defined sections of the data economy, such as the social media or search engine sectors,

quasi-monopolies are leading to suboptimal outcomes economically (Ducci 2020; Shapiro and Varian 2008). Digital platforms act as intermediaries between two or more user groups with interdependent demands in so-called two-sided markets (Veisdal 2020; Hagiu 2009; Boudreau and Jeppesen 2015). This is so because many platform providers need a critical mass of users to function properly, and in turn, the users draw more benefits from the platforms most in use.

Lock-in effects, such as access to a personal network or algorithmically personalized search results, make platform-switching costlier from the user's perspective. At the same time, the rising importance of digital services for the functioning of society means that some of these platform providers can now be considered part of a country's "critical infrastructure," just like telecommunications or electricity providers; they are no longer viewed as purely economic actors. While Europe clearly failed to ride the first wave of digital platform economies, which are mostly business-to-consumer (B2C) focused, whether European firms will be more successful in the next, more business-to-business (B2B) focused wave remains unclear. The peculiarities of data-driven digital markets give rise to the question of what the state can do to achieve and preserve digital sovereignty. In the following, we compare two distinct approaches.

4 DISCUSSION: TWO INTERNATIONAL POLITICAL ECONOMY APPROACHES

The different explanations for Europe's failure to create digital champions in the first wave of digital platform economies and the proposals for remedying this in the second wave essentially all try to answer the following question: What is the right balance between state intervention and market dynamics that offers the greatest added value for society in a globally networked data economy? In the following, we summarize the key notions of each of these economic policy approaches before proposing a way forward.

The **market liberal approach** proposed by Adam Smith and other (neo)classical economists contends that market dynamics in an open economy generate optimal outcomes. Accordingly, the state cannot enforce the emergence of digital champions. Even if this were possible, it would only prompt the replacement of foreign businesses with homegrown monopoly companies, which might enhance digital sovereignty on the state level but would do nothing to foster it on the individual or organizational level. The result of such political interventions can be observed in China, where politically imposed market entry barriers have created national digital champions pursuing business models comparable to those of US companies, leading to an even stronger market concentration than in the United States or Europe (Arsène 2015). In addition, any attempt to create European champions would risk decoupling the EU from cutting-edge developments in the rest of the world by limiting

the ability of foreign companies to operate in Europe. Indeed, foreign dominance in industries where such companies have a comparative advantage is not problematic from a market liberal perspective, as explained in the international division of labor theory by David Ricardo. This, however, does not say much about Europe's role in newly emerging technology areas. Rather than directly promoting companies, market liberalists focus on framework conditions. In the European context, they recommend completing the single market in digital services and goods to give European companies a bigger market and potential user base from which to scale up their businesses to global success. Restrictive regulations in some jurisdictions also hamper the growth of technology companies (Detrixhe 2018). Liberalists also note that financing for rapidly growing companies in Europe could be improved, for instance, by further integrating cross-border venture capital markets and providing tax incentives. Finally, labor mobility in the EU is much lower than in the US, making it more difficult for clusters of technology innovation and excellence such as Silicon Valley to emerge (Bauer and Erixon 2020).

In contrast, the **industry approach** regards (limited) state intervention and an active industrial policy as necessary to enable the rise of new technology clusters and to ensure a functioning market later on, especially in those areas of critical importance to a competitive and resilient economy and society (building on 19th century economist Friedrich List's infant industry argument). This line of thinking accepts short-term economic efficiency losses for the benefit of societal welfare in the longer term. According to this approach, unrestricted competition will result in market failure if it leads to monopolies or oligopolies dominated by foreign companies commanding vast amounts of data. The existence of higher market entry barriers in such industries compared to many other ones make it necessary for the state to help individual companies grow to a size where they can compete on a global level. Industrial policy advocates are concerned about the next technology wave of B2B digital platforms in the realm of the internet of things, as these technologies may affect many industries where Europe is traditionally strong (e.g., cars and machinery). Advocates of a more interventionist approach also maintain that even in the supposedly market-oriented US economy, many technology companies have benefitted from state intervention and public funding in R&D (Mazzucato 2011).

The industrial policy approach emphasizes the criticality of digital services and platforms for the state's digital sovereignty. According to this, theories of comparative advantage are inadequate to explain the complexity of today's digital economy, where control over data is just as important as economic efficiency gains (Carriere-Swallow and Haksar 2019). This would justify stronger regulation of the data economy, promoting data access and data-sharing (El-Dardiry, Dinkova, and Overvest 2021).

It is evident that, in recent years, the discourse on economic policy in Europe has shifted away from the market liberal approach and gone closer to an active industrial policy approach. This shift has only been augmented by the COVID-19 pandemic and the accelerated digital transformation, even as failed attempts to re-shore production of sensitive equipment have demonstrated the limits of state intervention and adverse effects of trying to upend global supply chains.

We conclude that a combination of elements of a market liberal and a more proactive industrial policy approach will be most conducive to the growth of European companies in the next wave of digital innovation. The liberal focus on framework conditions is useful in refocusing policy makers' efforts on deepening the incomplete EU single market, strengthening venture capital markets, and reminding policy makers of previous failed attempts of the state to pick individual firms and try to grow them into digital champions. On the other hand, according to an industrial policy approach, it may be advisable for the state to fund pilot projects in early-stage technology areas in partnership with industry to define new standards and rules for markets that are not yet fully consolidated. Policy makers can also regulate digital platform providers in such a way that they have to offer their customers different business relationships—for instance, providing free services in return for customer's personal data, or a subscription-model without data sharing, or even business models where users receive compensation for sharing more data. This would make the data economy more transparent and facilitate platform-switching. The draft EU legislation on Digital Markets and Digital Services already goes in that direction (European Commission 2020).

Importantly, when determining the right policy mix and balance between market openness and state intervention, policy makers should adopt a differentiated approach for different technology fields and stages of technology development. They may need to find the right moment to move away from an interventionist industrial policy and to a more market liberal approach. Even after the state has withdrawn from a more developed technology market (with higher technology readiness levels), it should maintain some oversight to prevent excessive market concentration. European policymakers should thus adopt bolder, more proactive industrial policies in early-stage technology areas and allow for more market competition in more advanced fields. A combination of liberal market-enabling measures and industrial policy regulation and standard-setting could shape a European "third way" to safeguarding digital sovereignty.

5 CONCLUSION AND OUTLOOK

This paper contributes to filling a gap in the debate on digital sovereignty by focusing on the level of the state and the geopolitical dimension of digital sovereignty. Using the example of the European debate on how to achieve and preserve digital sovereignty vis-à-vis global competition and

considering the market dynamics of the data economy, we adopt a political economy perspective that compares a market liberal and an industrial policy approach to answering this question.

Irrespective of the specific technology area, it is important for Europe to remain open to technology companies from abroad, so as to avoid creating a technosphere of its own that cannot access and benefit from innovation elsewhere. Instead, Europe should make its own model for the data economy attractive to the rest of the world by emphasizing trusted digital services and goods, data security and privacy. This way, digital sovereignty can be realized through *coopetiti*on (strategic cooperation and competition) globally.

The questions discussed in this paper can be investigated further. Research could build on the concept of digital sovereignty proposed here and review the political debate in other parts of the world, notably in the United States, China, and other parts of Asia. Furthermore, while this paper alluded to the consequences that different policy approaches on the state level may have on the level of organizations (businesses in particular) and individuals, further research could investigate the links between the state and these other two levels more thoroughly, and thus arrive at a more holistic understanding of digital sovereignty as a concept permeating all levels of society.

6 REFERENCES

1. Alt, R., Zimmermann, H. (2019). Electronic Markets on Platform Competition. Electronic markets, 29 (2), 143–149. https://doi.org/10.1007/s12525-019-00353-y.
2. Arsène, S. (2015). Internet Domain Names in China. China perspectives, 2015 (4): 25–34. https://doi.org/10.4000/chinaperspectives.6846.
3. Baik, J. (2020). Data Privacy Against Innovation or Against Discrimination? The Case of the California Consumer Privacy Act (CCPA). Telematics and Informatics, 52:101431. https://doi.org/10.1016/j.tele.2020.101431.
4. Bauer, M., Erixon, F. (2020). Europe's Quest for Technology Sovereignty: Opportunities and Pitfalls. ECIPE occasional paper 2020, 02. Brussels, Belgium. European Centre for International Political Economy. https://ecipe.org/wp-content/uploads/2020/05/ECI_20_OccPaper_02_2020_Technology_LY02.pdf.
5. Bendiek, A., Neyer, J. (2020). Europas Digitale Souveränität: Bedingungen und Herausforderungen Internationaler Politischer Handlungsfähigkeit. Demokratietheorie im Zeitalter der Frühdigitalisierung, edited by Oswald, M., Borucki, I. 103–25. Wiesbaden, Heidelberg: Springer VS.
6. Biersteker, T. (Ed.). (1999). State Sovereignty as Social Construct. Reprinting. Cambridge studies in international relations 46. Cambridge: Cambridge Univ. Press.
7. Biersteker, T. (2002). State, Sovereignty and Territory. Handbook of International Relations, 157–76. London, United Kingdom: SAGE Publications Ltd.
8. Boudreau, K., Jeppesen, L. B. (2015). Unpaid Crowd Complementors: The Platform Network Effect Mirage. Strat. Mgmt. J. 36 (12): 1761–77. https://doi.org/10.1002/smj.2324.
9. Bradford, A. (2020). The Brussels Effect: How the European Union Rules the World. New York, NY: Oxford University Press.
10. Braud, A., Fromentoux, G., Radier, B. and Grand, O. (2021). The Road to European Digital Sovereignty with Gaia-X and IDSA. IEEE Network 35 (2): 4–5. https://doi.org/10.1109/MNET.2021.9387709.
11. Brynjolfsson, E. Saunders, A. (2010). Wired for Innovation: How Information Technology Is Reshaping the Economy. Cambridge, Mass. MIT Press.
12. Carriere-Swallow, Y. Haksar, V. (2019). The Economics and Implications of Data: An Integrated Perspective. Departmental paper no. 19, 16. Washington, DC, USA: International Monetary Fund.
13. Couture, S., Toupin, S. (2019). What Does the Notion of "Sovereignty" Mean When Referring to the Digital? New Media & Society 21 (10): 2305–22. https://doi.org/10.1177/1461444819865984.
14. Detrixhe, J. (2018). Why Can't Europe Create Tech Giants Like the US and China? https://qz.com/1320983/why-arent-europes-technology-companies-as-big-as-in-the-us-and-china/.
15. Ducci, F. 2020. Natural Monopolies in Digital Platform Markets. Global competition law and economics policy. Cambridge: Cambridge University Press.
16. Edler, J., Blind, K., Frietsch, R., Kimpeler, S., Kroll, H., Lerch, C., Reiss, T. et al. (2020). Technologiesouveränität – Von der Forderung zum Konzept.

17. El-Dardiry, R. Dinkova, M., Overvest, B. (2021). Policy Options for the Data Economy: A Literature Review. CPB background document. The Hague: CPB Netherlands Bureau for Economic Policy Analysis. https://www.cpb.nl/sites/default/files/omnidownload/CPB-Background-Document-Policy-Options-Data-Economy-Literature-Review.pdf.

18. European Commission. 2020. Europe Fit for the Digital Age: Digital Platforms. https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2347.

19. Floridi, L. (2020). The Fight for Digital Sovereignty: What It Is, and Why It Matters, Especially for the EU. Philos. Technol. 33 (3): 369–78. https://doi.org/10.1007/s13347-020-00423-6.

20. Hagiu, A. (2009). Two-Sided Platforms: Product Variety and Pricing Structures. Journal of Economics & Management Strategy 18 (4): 1011–43. https://doi.org/10.1111/j.1530-9134.2009.00236.x.

21. Hummel, P., Braun, M., Tretter, M., Dabrock, P. (2021). Data Sovereignty: A Review. Big Data & Society 8 (1): 205395172098201. https://doi.org/10.1177/2053951720982012.

22. Katz, M. L. (1994). Systems Competition and Network Effects. The journal of economic perspectives.

23. Korff, B. S. A. (1923). The Problem of Sovereignty. Am Polit Sci Rev 17 (3): 404–14. https://doi.org/10.2307/1944043.

24. Krasner, S. D. (2012). Problematic Sovereignty: Contested Rules and Political Possibilities. New York: Columbia University Press. http://gbv.eblib.com/patron/FullRecord.aspx?p=909282.

25. Lambach, D. (2020). The Territorialization of Cyberspace*. International Studies Review 22 (3): 482–506. https://doi.org/10.1093/isr/viz022.

26. Mazzucato, M. (2011). The Entrepreneurial State: Debunking Public Vs. Private Sector Myths. London: Demos.

27. Pohle, J., Thiel, T. (2020). Digital Sovereignty. Internet Policy Review 9 (4). https://doi.org/10.14763/2020.4.1532.

28. Rochet, J.-C., Tirole, J. (2003). Platform Competition in Two-Sided Markets. *Journal of the European Economic Association* 1 (4): 990–1029. https://doi.org/10.1162/154247603322493212.

29. Rumana B., Heeks, R. (2017). Defining, Conceptualising and Measuring the Digital Economy: Development Informatics Working Paper No.68. Development Informatics Working Paper (68). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3431732.

30. Shapiro, C., Varian, H. R. (2008). Information Rules: A Strategic Guide to the Network Economy. 17. Boston, Mass. Harvard Business School Press.

31. Sirota, D. (2019). California's New Data Privacy Law Brings U.S. Closer to GDPR. TechCrunch, 2019. https://techcrunch.com/2019/11/14/californias-new-data-privacy-law-brings-u-s-closer-to-gdpr/?guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAACMQoZFpS9Cl7d5lj_PJzvHypRA4cvF3nKe6LbZ52Oe2P3Q41SpNww49FMJNLAuGD1a6GZE4p5rfloQI4uuxpXf3efNTJSXD5NsXVytFA1EYYephTds-PlUUmof-FAMT9w0WolhiJM4GL7pP1QFggfCZkKnYI9Zkdh4r1HTsXrLr&guccounter=2.

32. Syuntyurenko, O. V. (2015). The Digital Environment: The Trends and Risks of Development. Sci. Tech.Inf. Proc. 42 (1): 24–29. https://doi.org/10.3103/S0147688215010062.

33. van de Velde, E., Debergh, P., Wydra, S., Som, O. (2015). Key Enabling Technologies (KETs) Observatory: Second Report December 2015.

34. VDE. (2020). Technologische Souveränität: Vorschlag Einer Methodik Und Handlungsempfehlungen. VDE-Positionspapier.

35. Veisdal, J. (2020). The Dynamics of Entry for Digital Platforms in Two-Sided Markets: A Multi-Case Study. Electronic markets 30 (3): 539–56. https://doi.org/10.1007/s12525-020-00409-4.

36. Voss, W. G., Houser, K. A. (2019). Personal Data and the GDPR: Providing a Competitive Advantage for U.S. Companies. Am Bus Law J 56 (2): 287–344. https://doi.org/10.1111/ablj.12139.

37. Wittpahl, V., (Ed.) (2017). iit-Themenband - Digitale Souveränität: Bürger, Unternehmen, Staat. Berlin, Heidelberg: Springer Vieweg Open.

Proceedings of the Weizenbaum Conference 2022:
Practicing Sovereignty. Interventions for Open Digital Futures

I AM DISSOLVING INTO CATEGORIES AND LABELS — AGENCY AFFORDANCES FOR EMBEDDING AND PRACTICING DIGITAL SOVEREIGNTY

Pop Stefanija, Ana
imec-SMIT, Vrije Universiteit Brussel
Brussels, Belgium
ana.pop.stefanija@vub.be

Pierson, Jo
Hasselt University, Faculty School of Social Sciences
Vrije Universiteit Brussel, imec-SMIT
Hasselt/Brussels, Belgium
jo.pierson@vub.be

KEYWORDS

data sovereignty; algorithmic sovereignty; agency; autonomy; affordances; agency affordances

DOI: 10.34669/wi.cp/4.4

ABSTRACT

While the notion of digital sovereignty is loaded with a multitude of meanings referring to various actors, values and contexts, this paper is interested in how to actualize individual digital sovereignty. We do so by introducing the concept of agency affordances, which we see as a precondition for achieving digital sovereignty. We understand this notion as the ability to exercise power *to*, as autonomy and agency for (digital) self-sovereignty, and as *power over* the infrastructural sovereignty of the privately owned automated decision-making systems (ADM) systems of digital media platforms. Building our characterization of digital sovereignty on an empirical inquiry into individuals' requirements for agency, our analysis shows that digital sovereignty consists of two distinct but interrelated elements—data sovereignty and algorithmic sovereignty. Enabling practicable digital sovereignty through agency affordances, however, will require going beyond the just technical and extending towards the wider societal (infra)structures. We outline some initial steps on how to achieve that.

1 INTRODUCTION

"I believe that intrusions into someone's privacy affects their identity, personal development and even the process of becoming, being and remaining a person. Slowly, while working with my data I felt like I am dissolving in categories and labels." (Respondent 10)

The notion of sovereignty is defined in Merriam-Webster ("Definition of SOVEREIGNTY" n.d.) as "unlimited power over a country" but also "supreme power" and "freedom from external control." Looking at the synonyms section, we see the notions of "autonomy, freedom, independence, self-determination, self-governance" there. As Pohle and Thiel (2020) have pointed out, a shift is now happening; the meaning is moving away from the initial understanding of sovereignty as a condition to claim and exercise authority over a territory. While this meaning is still applicable, we can talk today about "the ability of individuals to take actions and decisions in a conscious, deliberate and independent manner."(Pohle and Thiel 2020, p. 11). When it comes to digital sovereignty, things aren't getting simpler. Digital sovereignty can be understood in many ways and can be used synonymously and interchangeably with related notions (e.g., data or cyber sovereignty)[4] and by different actors (Couture and Toupin 2019). The definition of the notion of (data) sovereignty also varies in terms of the actors involved, the contexts and domains referred to, and the values ascribed to it (Hummel et al. 2021). This ranges from governments understanding digital sovereignty as "the idea that states should reassert their authority over the internet" (Pohle and Thiel 2020, p. 2) to the claims of social movements that sovereignty relates to the "technologies developed from and for civil society" (Couture and Toupin 2019), to individual sovereignty (Pohle and Thiel 2020, Hummel et al. 2021).

This paper deals with the latter meaning of the notion and touches upon the notions of user (individual) autonomy, self-determination, control, and agency. We define digital sovereignty as an individual's ability to have control over their data and digital "life" and an ability to reject, oppose, and steer their own behavior with self-determination and autonomy, freed from external influences. This control and autonomy over one's digital life also extends to the outputs of the data processing activities (profiling, personalization, recommendations, and automated decision-making) and implies ability and the capacity to act with autonomy, control, and self-determination based on self-reflection when facing or being subjected to these algorithmic decisions. We understand this notion as an ability to exercise power *to*—as autonomy and agency for (digital) self-determination, self-actualization, and self-sovereignty—and as a *power over* the infrastructural sovereignty of the private owned ADM

[4] For a good overview of the notion of digital sovereignty and data sovereignty, see Pohle and Thiel (2020), Couture and Topin (2019), and Hummel et al. (2021).

systems of digital media platforms. The results of the empirical research we conducted show that digital sovereignty consists of two inter-related forms of sovereignty—what we can call *data sovereignty* and *algorithmic sovereignty*. The analysis also shows that sovereignty is closely intertwined with the notion of agency. To enable the practicing of digital sovereignty, we propose enabling sovereignty through agency affordances, understood as programmed functions and embedded features in the algorithmic systems that should enable, afford, and make the agency of individuals operational and actionable (authors, in preparation).

With this paper, we present the results from our empirical research and elaborate on the characteristics of digital sovereignty and its two elements—data and algorithmic sovereignty. We then introduce the notion of agency affordances and propose ways for their embedding in technology. We further discuss what else is needed to practice digital sovereignty.

2 METHODOLOGICAL APPROACH

Our theorizing about individuals' digital sovereignty and its relationship with agency is the result of an empirical inquiry aiming to discover the requirements of individuals for agency when they are interacting with or subject to automated decision-making (ADM). With our research design, we aimed to obtain empirical insights from "real-life experiences." That meant eliciting insights based on a real interaction between the participants and the ADM system of their choice and asking them to formulate their concerns and requirements after a period of interaction and reflection. To capture this, we developed a structured diary, where participants recorded the process and noted down their expectations, experiences, and needs and requirements. We opted for the diary method because if "facilitates critical knowledge, and involves reflexivity" (Fisher, 2020, p. 2), resulting in individuals gaining new knowledge about themselves but also the ADM systems. As such, this is a knowledge that would not have been accessible otherwise (ibid.). The diaries (structured in 15 questions) recorded the experiences of participants in interacting with a platform of their choice and captured their reflections on that interaction. Based on this reflective exercise, participants were able to formulate and voice their expectations, needs, and requirements regarding agency and trust. The notions of data and algorithmic sovereignty emerged from the analysis of these requirements for agency. As we will elaborate, agency—or the ability to act—is a crucial precondition for realizing and practicing digital sovereignty. Yet, agency in the context of interaction with digital media platforms must be accompanied by a few elements and conditions to be met.

3 METHODOLOGICAL SETUP

The research took place over a three-month period, October to December 2020. It included 47 participants, who were students at an international graduate program in Belgium. Participants could choose from eight platforms (Facebook, Google, Twitter, Instagram, Tinder, Spotify, Netflix, TikTok) as a platform they would like to interact with. They were provided with a template for the subject access request (SAR) (Veale 2019). The research design consisted of a multi-stage process, including the completion of a survey, the submission of an SAR, and purposeful interaction with the tools the platforms themselves offer for data collected and held about the participants (we refer to these as platform transparency tools). At the end, we were provided by each of the participants with the following outputs: filled-in diaries, lists of all the inferences/categories assigned to them by the platform of their choice, an illustration of their real identity, an illustration of the algorithmic identity assigned by the platform (their reading of), and a list of requirements for agency and for trust. The authors guided the participants for the entire duration of the process.

4 WHAT KIND OF DIGITAL SOVEREIGNTY?

Our analysis was predominantly focused on the requirements of the participants to have more agency when interacting with platforms' ADM systems. We used the diaries in their entirety to obtain a general feel, but in particular, we focused on the explicitly stated requirements for agency. In total, there were 159 unique requests by the 47 participants. Through a process of iterative and inducive coding, using the framework for thematic analysis (Braun and Clarke 2006), we extrapolated three interrelated requirements for agency: the *ability to see* (transparency), the *ability to know* (explainability), and the *ability to act*. We will describe them here briefly, because they are relevant for the consequent discussion of digital sovereignty.

The *ability to see* pertains to the requirement to be provided with information—that is, to be presented with or provided access to a variety of information. This information should provide insights into the relation of the individual to the particular system, including, among others, any actors that receive, collect, or use the data. The requests related both to *data provenance & cycle transparency* and to *information structuring and access.* These concern data origin and usage—whom their data is shared with, what is considered data, what/who is the source of the data, what data is collected, how and where is it stored, and so on—as well as requests to make the information more visible and/or information-structuring simpler and understandable.

The *ability to know* is related to the ability to engage in sense-making and understanding; it is about making things *knowable* and about being provided with the opportunities and tools to

understand and comprehend. Our data show that the most frequent request is to be able to acquire knowledge on how automated decisions are made, followed by requests for knowledge about how inferences about participants are made.

The *ability to act* concerns concrete requests related to the ability of an individual to make an autonomous and authentic decision about and for themselves and act upon it. It is related to the entwined notions of control, autonomy, and sovereignty. It concerns all the elements of the AI/ADM system—the data, the system itself, and sometimes even the business model. This *ability to act* is dependent both on the *ability to see* (information and transparency) and the *ability to know* (explainability and understanding). While the former is not enough in itself to guarantee the latter, the possibility to have information and to understand, reflect, and decide based on that, is fundamental.

In this paper we focus on this third requirement, the ability to act, because it is here that we see the visible prominence of the notion of digital sovereignty. Our analysis (Figure 1) showed that needs for *abilities to act* are predominantly related to two distinct but interrelated elements and processes: the data (cycle) and the (automated) decision-making. This corresponds to what we call *data sovereignty* and *algorithmic sovereignty*.

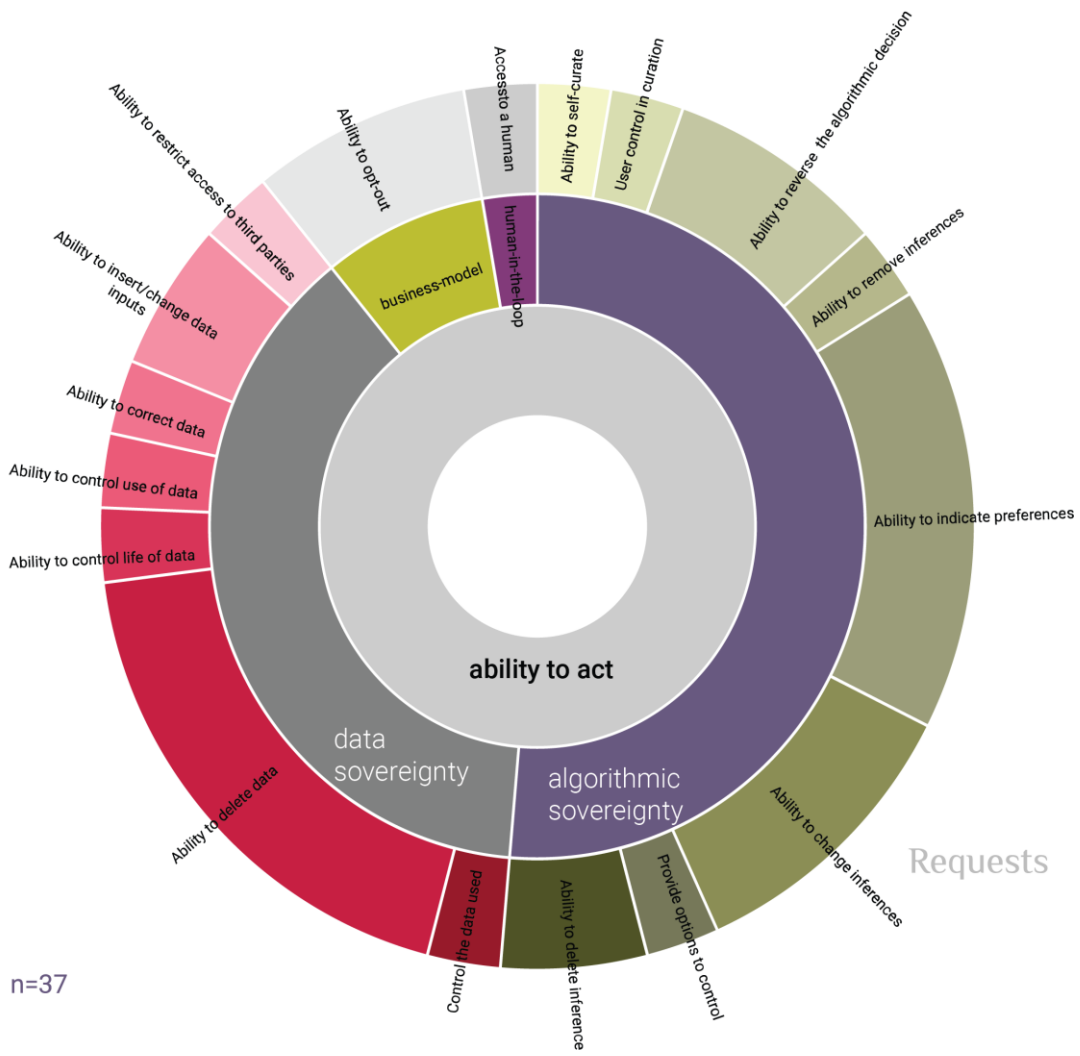


Figure 3. Requests for ability to act. The coded requests are grouped and color-coded per type of sovereignty and frequency.

4.1 DATA SOVEREIGNTY OR THE ABILITY TO ACT IN RELATION TO DATA

What our respondents' requirements show is that they request the ability to act in relation to two distinct data processes—data provenance and the data cycle. When it comes to data provenance, they want the ability to act on what data is collected, how, and by whom. The data cycle requests pertain to the possibility to act/decide on how data is shared, with whom it is shared, what it is used for, and how it is used, in all phases of the data cycle—data design, data capture, data processing, and data usage.

As is evident from Figure 2, the ability to delete data is one of the most prominent requests—"users should be able to fully delete the trace on themselves, not just photos and their account, but also the data that was built up by the platforms themselves to then use it for profiling purposes"

(Respondent 14) as well as data shared with third parties—"… the user should be able to completely delete his page from the social network and all data that was transferred to third parties should be automatically deleted" (Respondent 31). The ability to have/retain control seems to be an almost equally important request—users wanted to have the ability to control the use and life of data once it has been collected. This encompasses the requests to be able to change, modify, delete, and opt-out.

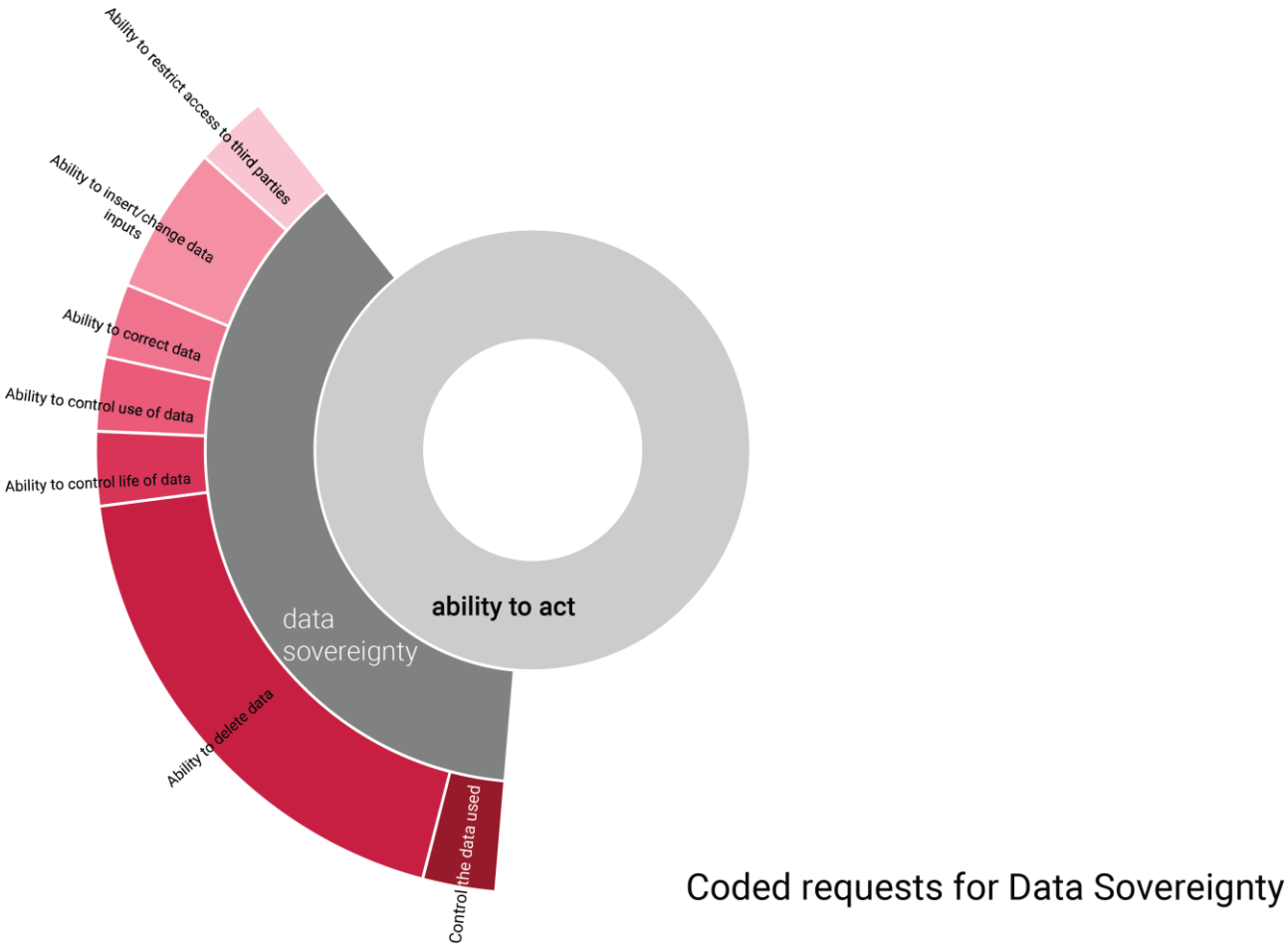


Figure 2. Requests for data sovereignty.

What we see from our respondents is a clear description of what Hummel et al. (2018, see also 2021) call data sovereignty—"meaningful control, ownership, and other claims to data or data infrastructures." (Hummel et al. 2018, p. 12). This sovereignty is related to the ability to "steer data flows and/or to govern informational resources." (Hummel et al. 2018, p. 10) and as such is related closely to control, power and autonomy. As is evident in the figure above, control—regarding different aspects of data collection, processing, and use—features very prominently. The ability to take control over these processes implies having *power over* datafication systems and the *power to* remedy power imbalances between users and the systems "processing" them. Being able to articulate and enforce claims of power about their data and being aware of the flow of their personal data reverses the power roles and imbalances and gives individuals the ability for reflexivity, agency, and

autonomy. This power also implies the ability to set up privacy boundaries by constraining access to data (e.g., to third parties), but also to steer and govern informational resources (an important element for data sovereignty, as outlined by Hummel et al. 2018). The element of autonomy—represented as an ability to correct and/or delete data inputs or to control the purpose for which data may be used—refers to the ability to act authentically, according to one's wishes and needs, without interference from external parties. This autonomy, sovereignty, and authenticity implies the opportunity and ability to also challenge, oppose, and reject.

4.2 ALGORITHMIC SOVEREIGNTY OR THE ABILITY TO ACT IN RELATION TO AUTOMATED DECISION-MAKING

The requests related to decision-making are related to the outputs of the system and the ability of individuals to have an active role and agency when they are subjected to the workings of the ADM/AI systems. We refer to this type of sovereignty as *algorithmic sovereignty*. As the analysis shows, this request for sovereignty is predominantly related to the *ability to indicate preferences* and the various abilities related to the inferences. The requests to remove, change, and delete inferences are requests for more control and autonomy over the digital/algorithmic self that is used to produce algorithmic outputs. It is the ability of the user to impose their own way of seeing themselves, their algorithmic identity, and ultimately their authenticity. This is a request for authenticity, a desire to be seen as one identifies oneself. And this is, most often, the opposite of what the system, from a power perspective, ascribes to them and formulates as their interests, wishes, and needs. We can see this also from the diary entries, where respondents had to depict their own sense of identity and the algorithmic identity assigned to them by the platforms according to the inferences constructed about them. As one respondent said, the real them is a "diverse person with many different faces" but the algorithmic them is a "human taken apart into data, a snapshot in time that gets algorithmically exploited."

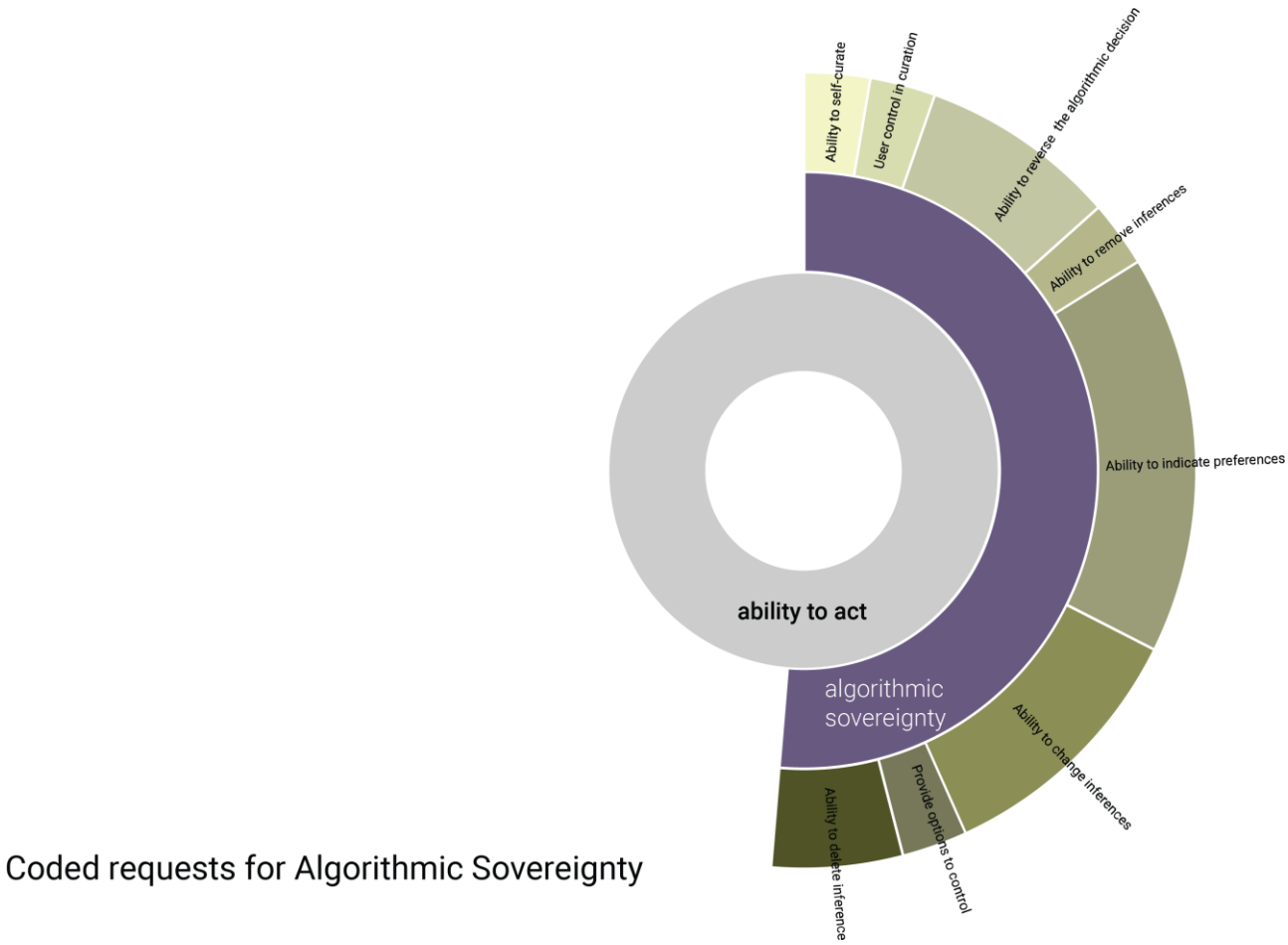


Figure 3. Requests for algorithmic sovereignty.

This request for authenticity is closely related to the request for self-determination, which is seen as an ability to reverse an algorithmic decision. Since most of the ADM outputs, especially in the case of social media platforms, are based on algorithmic profiling, the autonomy and sovereignty are seen as a possibility to disagree, reject, and actively change the outcomes of the algorithmic process. This is, in essence, a request to be freed from algorithmic governmentality—from the steering of subjectivities and life chances based on datafication, classifications, ranking, sorting, and predicting—and a request to be able to actively govern their own lives.

5 AGENCY AFFORDANCES FOR DIGITAL SOVEREIGNTY

To make these data and algorithmic sovereignty requests possible, we propose the notion of *agency affordances*. In conceptualizing this notion, we build on the requirements that our participants

elaborated in their diaries and the list of requirements they provided, as well as from agency theory[5] and affordances theory[6].

As we elaborate elsewhere (authors, in preparation), we define agency affordances as functions that are, first, *programmed* and embedded at an infrastructure level that should allow and encourage the actualization of agency. This occurs, second, through features and elements made visible and *promoted* at an interface level, coupling the possibilities for action (to act) with the ability to act. However, agency affordances are non-determining, relational, dynamic, context-dependent, situated, and come in gradations and variable forms. They should ensure the possibility and actualization of control, autonomy, authenticity, and ultimately sovereignty, leading to true (digital) empowerment of individuals. It should enable individuals to become sovereigns on their own, without external inferences steering their behavior and impacting life chances. They should ensure the possibility to both enable and enforce self-data governance needs and power for self-governance. This will help remedy power imbalances and facilitate, as much as possible, a shift from platforms as sovereigns to individuals as self-sovereign.

6 EMBEDDING PRACTICABLE DIGITAL SOVEREIGNTY

How do we acquire digital sovereignty both as *power to* and as autonomy and agency for (digital) self-determination, self-actualization, and self-sovereignty? For the individuals to be able to practice digital sovereignty, it should be enabled, as an *ability*, through the embedding of agency affordances withing the technological system. These affordances should be introduced as functions at infrastructure level and as features at interface level. They should enable individuals to exercise *power over* the infrastructural sovereignty of the ADM systems of digital media platforms. This will require careful architecture and design planning when setting up these systems or when modifying them to respond to the requests for digital sovereignty. When doing so, the relevant actors should also account for the different contexts, the different individual skills, the different needs, and the different abilities to understand; in that sense, the process should be highly contextual and take into consideration the various educational, social, cultural, and similar environments and circumstances of the individuals.

However, the lived experience of actualizing and practicing digital sovereignty is not just a matter of technology and extends to encompassing the wider societal (infra)structures. That would mean enabling and securing digital sovereignty by introducing dynamics and conditions that enable

[5] Based on the texts by the following authors: Neff and Nagy, 2016; Neff et el. 2012; Nagy and Neff 2015; Couldry 2014; Feenberg 2011; Kennedy et al. 2015; Lorusso 2021; Milan 2018.
[6] Based on Davis and Chouinard 2017; Davis 2020; Neff and Nagy 2016; Nagy and Neff, 2015; Bucher and Halmond, 2018; Evans et al. 2017; Dahlman et al. 2021; Neff et al 2012, Hutchby, 2001.

and demand sovereignty via regulations, institutions, and organizations. This should make sure that digital sovereignty is also a *right*, "something towards which we should aim" (Hummel et al. 2021, 13) and not just an ability. The processes for setting the institutional norms for embedding agency affordances and thus securing and enforcing digital sovereignty via regulations, regulatory bodies, and (informal)institutions could range from introducing agency affordances as mandatory by law, to undertaking literacy initiatives, to introducing what Wachter and Mittelstadt (2019) suggest—introducing a new right, the *right to reasonable inferences* "by which meaningful control and choice (p. 13) over inferences and profiles are granted to data subjects." (p. 14).

While individuals are being entangled in a net of powerful data and digital sovereigns that do not just control the entire digital infrastructure but also steer individuals' lives and impact their life chances, the embedding of agency affordances as a way of practicing digital sovereignty might not be the easiest option. But it is a tangible way to impose "empowerment by design" (Pierson, 2022) and ultimately sovereignty by design and default.

7 ACKNOWLEDGMENTS

The research was done as part of the project DELICIOS "An integrated approach to study the delegation of conflict-of-interest decisions to autonomous agents" (G054919N), funded by the Fonds voor Wetenschappelijk Onderzoek – Vlaanderen (FWO).

8 REFERENCES

1. Braun, Virginia, and Victoria Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3 (2): 77–101. https://doi.org/10.1191/1478088706qp063oa.
2. Bucher, T., and A. Helmond. 2018. "The Affordances of Social Media Platforms." In *The SAGE Handbook of Social Media*, 233–53. Sage Publications. https://dare.uva.nl/search?identifier=149a9089-49a4-454c-b935-a6ea7f2d8986.
3. Couldry, Nick. 2014. "Inaugural: A Necessary Disenchantment: Myth, Agency and Injustice in a Digital World." *The Sociological Review* 62 (4): 880–97. https://doi.org/10.1111/1467-954X.12158.
4. Couture, Stephane, and Sophie Toupin. 2019. "What Does the Notion of 'Sovereignty' Mean When Referring to the Digital?" *New Media & Society* 21 (10): 2305–22. https://doi.org/10.1177/1461444819865984.
5. Dahlman, Sara, Ib T Gulbrandsen, and Sine N Just. 2021. "Algorithms as Organizational Figuration: The Sociotechnical Arrangements of a Fintech Start-Up." *Big Data & Society* 8 (1): 20539517211026704. https://doi.org/10.1177/20539517211026702.
6. Davis, Jenny L. 2020. *How Artifacts Afford: The Power and Politics of Everyday Things*. Design Thinking, Design Theory. Cambridge, MA, USA: MIT Press.
7. Davis, Jenny L., and James B. Chouinard. 2017. "Theorizing Affordances: From Request to Refuse:" *Bulletin of Science, Technology & Society* 36 (4): 241–48. https://doi.org/10.1177/0270467617714944.
8. "Definition of SOVEREIGNTY." n.d. In Merriam-Webster. Accessed February 28, 2022. https://www.merriam-webster.com/dictionary/sovereignty.
9. Evans, Sandra K., Katy E. Pearce, Jessica Vitak, and Jeffrey W. Treem. 2017. "Explicating Affordances: A Conceptual Framework for Understanding Affordances in Communication Research." *Journal of Computer-Mediated Communication* 22 (1): 35–52. https://doi.org/10.1111/jcc4.12180.
10. Feenberg, Andrew. 2017. "Agency and Citizenship in a Technological Society." In *Spaces for the Future*, 1st ed., 98–107. Routledge. https://doi.org/10.4324/9780203735657-10.
11. Fisher, Eran. 2020. "The Ledger and the Diary: Algorithmic Knowledge and Subjectivity." *Continuum* 34 (3): 378–97. https://doi.org/10.1080/10304312.2020.1717445.
12. Herlo, Bianca, Daniel Irrgang, Gesche Joost, and Andreas Unteidig, eds. 2022. *Practicing Sovereignty*. Architecture and Design. Bielefeld, Germany: Transcript. https://www.transcript-publishing.com/978-3-8376-5760-9/practicing-sovereignty/.
13. Hummel, Patrik, Matthias Braun, Steffen Augsberg, and Peter Dabrock. 2018. "Sovereignty and Data Sharing." *ITU Journal: ICT Discoveries* 1 (2). https://www.itu.int:443/en/journal/002/Pages/11.aspx.
14. Hummel, Patrik, Matthias Braun, Max Tretter, and Peter Dabrock. 2021. "Data Sovereignty: A Review." *Big Data & Society* 8 (1): 2053951720982012. https://doi.org/10.1177/2053951720982012.
15. Hutchby, Ian. 2001. "Technologies, Texts and Affordances." *Sociology* 35 (2): 441–56. https://doi.org/10.1017/S0038038501000219.

16. Kennedy, Helen, Thomas Poell, and Jose van Dijck. 2015. “Data and Agency.” *Big Data & Society* 2 (2): 2053951715621569. https://doi.org/10.1177/2053951715621569.

17. Lorusso, Silvio. 2021. “The User Condition: Computer Agency and Behaviour.” December 2, 2021. https://theusercondition.computer/.

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

19. Nagy, Peter, and Gina Neff. 2015. “Imagined Affordance: Reconstructing a Keyword for Communication Theory.” *Social Media + Society* 1 (2): 1–9. https://doi.org/10.1177/2056305115603385.

20. Neff, Gina, Tim Jordan, Joshua McVeigh-Schultz, and Tarleton Gillespie. 2012. “Affordances, Technical Agency, and the Politics of Technologies of Cultural Production.” *Journal of Broadcasting & Electronic Media* 56 (2): 299–313. https://doi.org/10.1080/08838151.2012.678520.

21. Neff, Gina, and Peter Nagy. 2016. “Talking to Bots: Symbiotic Agency and the Case of Tay.” *International Journal of Communication* 10: 4915–31.

22. Pierson, J. (2022). Media and Communication Studies, Privacy and Public Values: Future Challenges, in: González-Fuster, Gloria, van Brakel, Rosamunde and De Hert, Paul (eds.) *Research Handbook on Privacy and Data Protection Law: Values, Norms and Global Politics*, Cheltenham: Edward Elgar Publishing: 175-195.

23. Pohle, Julia, and Thorsten Thiel. 2020. “Digital Sovereignty.” *Internet Policy Review* 9 (4). https://policyreview.info/concepts/digital-sovereignty.

24. Taylor, Linnet. 2021. “Public Actors Without Public Values: Legitimacy, Domination and the Regulation of the Technology Sector.” *Philosophy & Technology* 34 (4): 897–922. https://doi.org/10.1007/s13347-020-00441-4.

25. Veale, Michael. 2019. “A Better Data Access Request Template.” https://michae.lv/access-template/.

Proceedings of the Weizenbaum Conference 2022:
Practicing Sovereignty. Interventions for Open Digital Futures

MACHINE LEARNING AND THE END OF THEORY

REFLECTIONS ON A DATA-DRIVEN CONCEPTION OF HEALTH

Guersenzvaig, Ariel
Elisava Barcelona School of Design and Engineering, UVic-UCC
Barcelona, Spain
aguersenzvaig@elisava.net

KEYWORDS

machine learning, health, theory, normative concepts

DOI: 10.34669/wi.cp/4.5

ABSTRACT

Taking the notion of health as a *leitmotif*, this paper discusses some conceptual boundaries for using machine learning—a data-driven, statistical, and computational technique in the field of artificial intelligence—for epistemic purposes and for generating knowledge about the world based solely on the statistical correlations found in data (i.e., the "End of Theory" view). The thrust of the argument is that prior theoretical conceptions, subjectivity, and values would—because of their normative power— inevitably blight any effort at knowledge-making that seeks to be *exclusively* driven by data and nothing else. The conclusion suggests that machine learning will neither resolve nor mitigate the serious internal contradictions found in the "biostatistical theory" of health—the most well-discussed data-driven theory of health. The definition of notions such as these is an ongoing and fraught societal dialogue where the discussion is not only about what *is*, but also about what *should be*. This dialogical engagement is a question of ethics and politics and not one of mathematics.

1 INTRODUCTION

An influential argument in favor of using artificial intelligence (AI) for epistemic purposes can be found in Chris Anderson's essay "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete" (Anderson, 2008), where he argues that big data and the AI tools used to process them offer a new way of understanding the world based on the statistical correlations between data. Correlations make causal explanations —i.e., human-made (conceptual) causal models and theories—unnecessary for scientific progress. Anderson does not just propose the use of AI to computationally support scientific discovery and theory generation; he wants AI to take the lead because "science can advance even without coherent models." Data-driven discovery is also defended by the astrophysicist Kevin Schawinski: "Let's erase everything we know about astrophysics. To what degree could we rediscover that knowledge, just using the data itself?" (Cited in Falk, 2019). His "generative" approach represents a much weaker—yet more plausible—version of Anderson's argument. Schawinski (et al., 2018) concedes that human insight is still required for high-level interpretation, which enables an expert to make sense of the discoveries. For some, an instantiation of this perspective can be found in the case of *AlphaFold*, an AI system that has been able to accurately predict the 3D structure of a protein, thus solving one of the great contemporary challenges of biology (Heaven, 2020).

For space reasons, I will not explore the view in detail nor the various epistemological questions that emerge from it (for a detailed treatment, see, e.g., Casacuberta and Vallverdú, 2014). Rather, I shall engage with the thrust of the argument—very succinctly laid out in the previous paragraph—indirectly and try to scrutinize whether data makes theories and previous conceptions truly redundant. To structure the discussion, I will examine whether I can use machine learning (ML)—possibly AI's most popular technique nowadays—to resolve or at least mitigate some of the internal contradictions found in a well-discussed data-driven theory of health that seeks to define what health is—the "biostatistical theory" of health—which I will first briefly introduced below.

2 CONCEPTS OF HEALTH AND TELLING WHO'S HEALTHY

Before jumping to machine learning, I will summarily consider some relevant aspects around the notion of "health," which will be the *leitmotif* here. First and foremost: there is no consensus on what health is. In the Western literature on health, we find, on the one hand, "naturalist" theories, whereby health is a value-free notion that is determined by empirical facts. On the other, we find "normativist" theories, whereby health is essentially value laden. I will briefly outline the naturalist view, which pursues a descriptive goal like Anderson's: to derive knowledge from statistical data.

Possibly the most vigorously debated naturalist perspective on health is the "biostatistical theory" proposed by Christopher Boorse (1977; 2014). This theory rests on a nonnormative understanding of biological function and a statistical notion of the concept of "normality." For Boorse, health and disease are nothing more than biological states. In this sense, to say that an organism is healthy is to describe a natural fact and not to make an assessment of it in terms of good or bad, desirable or undesirable, and so on. Boorse states "if diseases are deviations from the species biological design, their recognition is a matter of natural science, not evaluative decision" (1977, p. 543).

As will be made clear shortly, Boorse's biostatistical theory fits nicely with the end of theory proposed by Anderson. Boorse (1977, p. 542) maintained that health is the "statistical normality of function" and that "the normal is the natural" (1977, p. 554). Diseases are "internal states that depress a functional ability below species-typical level" (1977, p. 542; 2014, p. 684). An organism is thus healthy when its functioning conforms to its natural design and function. Boorse's theory is much richer than I can cover here, alas, yet the upshot is that health is the fitness of an organism to perform its normal functions with statistically normal efficiency under typical conditions.

Typical levels for a species are those close to the statistical mean (Boorse, 1977, pp. 558–559). Although "normal" levels could be determined statistically for the whole species, from a clinical perspective, it would be impossible to conduct a comparison at a species level. Hence, a smaller reference class is needed. Since species design seems to be contingent on sex, age, and race, the statistical abstractions should be made from reference classes smaller than species (Boorse, 1977, p. 558). To assess the normality of a biological state for a subgroup within a species, Boorse needs some sort of benchmark of normality. To determine whether a particular organism is healthy in relation to the species-typical level Boorse introduces the notion of a "reference class."

A reference class is "a natural class of organisms of uniform functional design; specifically, an age group of a sex of a species" (1977, p. 555). Examples of reference classes would be "a 35-year-old white woman" or "a neonate of Aymara ancestry." In short, according to Boorse, if we want to establish the health of a neonate's heart, we should compare it to the hearts of other neonates, factoring in sex and race, and not to an average adult human heart, as an adult with the constant heart rate of a neonate would be considered diseased, and vice versa.

However reasonable and clinically necessary reference classes may be, Boorse undermines himself methodologically by introducing them—and rather evidently so. This is an objection noted by Elselijn Kingma (2010): It is not clear *why* it would be appropriate from a naturalist, nonnormative perspective to factor sex, age, and race in when calculating normality and not other criteria. There are no empirical facts that determine that "neonates" represent an appropriate reference class, but "people

with beards" or "children with dental cavities" are not. Indeed, both beards and caries are statistically frequent. What's more, even allowing for sex to be partially constituted by some empirical indicators, such as testosterone levels, its status as a full-blown natural category has been a hotly debated issue since the 1990s (see Butler, 1990).

Kingma convincingly shows that Boorse cannot justify his choice of appropriate reference classes without involving value judgements and prior theories and conceptions of health. If Boorse's theory seeks to stand independently of normative knowledge, it should be able to offer a value-free explanation of which criteria constitute an appropriate reference class. It is not enough to assert that "sex," race," and "age" are (the) appropriate reference classes. In other words, for the biostatistical theory to be truly naturalistic, the required reference classes must be determined and justified neutrally and empirically objectively without underlying value judgements. And for his critics, this is what Boorse's biostatistical theory fails to achieve.

Would it be possible to use ML to "end the theory" that is inherent in the above-mentioned reference classes? Could ML release the biostatistical theory from the insidious values, conceptual models, and theories that sabotage its quest for nonnormativity? If we succeed in this task, Anderson's views on the end of theory would become more compelling and Boorse's work would be free from its internal methodological contradictions. Above all, a truly naturalistic theory of health would be closer to hand.

3 MACHINE LEARNING TO THE RESCUE?

Why "sex," "age," and "race" instead of other criteria? Kingma asked. Fortunately, given the possibilities of machine learning, we could virtually limitlessly extend the range of reference classes beyond these three. Certainly, an ML system could use any attribute of the human body that can be incorporated into database tables: from eye color to bone density to hair thickness to lung capacity to weight. For instance, an ML system could then be trained with anthropometric data: skull shape and volumetric measurements, abdominal circumference, limb alignment, eye color, and so on. To train the system, we would first need to label the input data so that the system could develop a model from it and assign an output label for a new value, i.e., a result in terms of "healthy" or "diseased."

Alas, this would not satisfactorily address Kingma's objection regarding prior normativity and subjectivity, which would still be detected in the defining training variables. "Why are *these* signs used and not others?," we might ask. Why skull shape or eye color? The only difference is that instead of having three criteria without atheoretical justification, we would possibly have many more.

Perhaps there might be a way out of this tangle. Given the sheer number of reference classes that could be defined, it would be conceivable to make health assessments by considering randomly

chosen classes. If we were to work with a reference class based on *multiple* data sources, this could perhaps bring us closer to assessing an individual's health status in a non-normative way.

This seems technically viable thanks to "random sampling." The system could be trained with anthropometric data, medical history data, and clinical signs data from people labelled as healthy or diseased. In this case, though, we would only use a fraction of the available data, randomly selecting which database tables are considered or left out in the construction of the model. So, the system could include data about the swelling (or lack thereof) of the lymph nodes but leave out data about blood pressure or head circumference.

However, this does not remove the normative influence. Indeed, while the choice of classes to be used as a benchmark would be random, the pool of reference classes the system could choose from would be normatively and theoretically justified. The very choice to use anthropometric and clinical database tables is itself a decision that is based on the value judgements and prior theories that underpin the judgement about what database tables to include. And this does not bring us any closer to having to accept that prior theory has become unnecessary thanks to machine learning.

Still, a defender of the end-of-theory view might retort: If the problem is in the choosing, what if the system could be trained with available data of *any* kind? Perhaps nonmedical data, which is prima facie neutral about health (e.g., high school grades, social network activity, data from tax returns, parking violations records, etc.) could be used. All that would be necessary is to train the system with a dataset containing the nonmedical records of healthy and diseased groups or individuals. After a while, the system would detect salient features in the data and identify connections between the medical and the nonmedical data.

However, another evident problem emerges, that of circularity. In supervised machine learning, the putatively neutral data needs to be connected to health (or to a proxy thereof) to be able to generate a result. In the same way that a bird-identification app needs to be taught (through labels) what different types of birds look like to be able to assign a label to an image of a new bird, the health system would still need to find patterns in the parking records or in the high-school grades belonging to healthy or diseased individuals or groups. Yet to train the system in this manner, we would necessarily require a prior conception of health and disease to label the data. It is precisely this that enables individuals or groups to be classified as either healthy or diseased! And this manifestly violates the very theory-free approach we are trying to achieve.

Undoubtedly, supervised machine learning can be used to make valuable assessments of health based on large volumes of data once the appropriate reference classes have been defined and the data has been labelled, but it is far from making prior theory redundant. On the contrary, it highlights how data is intertwined with theory.

4 THEORY PRECEDES DATA

A defender of the end-of-theory position might claim that unsupervised machine learning is the way to go as, it would not be tainted by circularity. Indeed, unsupervised ML does not need labelling up front, and the reference classes could emerge as clusters from the data alone thanks to correlations. This is a point also made by Anderson (2008), for whom big data "allow us to say: 'Correlation is enough.' We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot."

It is certainly possible to cluster data without labels, for instance, by using profiling to detect patterns or structures present in the data that have not been previously hypothesized. Through techniques such as profiling, classifications (i.e., clusters) can be made without the need for causal models or other theoretical explanations. What's more, the system could deal with vast purely numerical vectors, whereby the original attributes—i.e., the column labels—in the database would not even need to be explicit at all. Only the numbers would be required.

We could generate reference classes thusly. Yet, some questions remain. Would these clusters truly *precede* theory and models? Would the classes be naturalist in the strict sense? To answer these questions, the discussion needs to turn to the nature of data itself. Since the issue is vast, I will be content to outline reasonable doubts about the possibility of unsupervised ML being able to generate naturalist, *atheoretical* reference classes.

The first aspect to bear in mind is that data are not directly and neutrally incorporated into systems as if they were a mirror of empirical reality. Data need to be collected and processed in order to be computationally readable. This first step already implies a reduction of the complexity of the world to a few database fields. This reduction, contrary to the intentions of naturalists, is marked by values such as efficiency, effectiveness, cost-effectiveness, budget limitations, and so on. No researcher simply "throws numbers" into a computer system.

Bowker and Star (2000) famously showed how classification systems shape and are themselves shaped by perspectives on the world and by social interactions. Data are not a Platonic entity. Data are a construct that is *made appropriate* to the systems and classification schemes in which they are incorporated according to some goals or purposes. Categories and attributes make some aspects visible while making others invisible. They are never a mere naturalist reflection of reality.

There is a second problem: misrepresentation in data selection ("sampling bias"), a common problem in datasets. An example is found in artificial intelligence systems that aim to assist

dermatologists in the detection of skin cancer. These systems exhibit great potential by achieving levels of prediction comparable or superior to that of dermatologists (Esteva et al., 2017; Fink et al., 2020). However, one grave problem from the perspective of justice is that they are much more accurate with light skin than with dark skin, which likely has to do with the datasets employed to train the systems to recognize potential moles (Adamson and Smith, 2018).

So, we see that approaches that start from data might yield reference classes, but there is no theory-free way of assessing the validity of these classes to determine that these are adequate from the statistical, clinical, and justice perspectives. We necessarily need auxiliary theories to align the three perspectives, starting from the very beginning at the data collection stage.

Relatedly, there is a third obstacle in the road to naturalist, theory-free reference classes: Artificial intelligence systems are characteristically affected by structural biases that go beyond sampling bias. Consider the gender bias that not only plagues medical data but medicine itself. Its history shows a *structural* lack of interest in women's health. Let's review a few examples. Eight of the 10 prescription drugs that were withdrawn from the US market in the period 1997–2001 posed greater health risks to women than to men (USGAO, 2001). Diseases are ignored when they do not affect men, as in the case of endometriosis (Huntington and Gilmour, 2005). Procedures and therapies might have distinct effects on men and women, yet this can go unnoticed for many years until women are included in controlled trials (Ridker et al., 2005). As happened with COVID-19 vaccines, the effects of medical interventions on menstruation seem to be an afterthought. Indeed, changes to period patterns and vaginal bleeding are not included among the common side effects of COVID-19 vaccination listed by the UK's regulatory agency MHRA, yet these events are reported to be frequent shortly after vaccination (Male, 2021).

In short, bias in machine learning is first and foremost a matter of justice and structural inequalities; it is not only a technical issue of statistical representativeness. There is a vast literature related to how race, gender, age, educational level, cognitive abilities, and many other vectors of unfairness (e.g., Benjamin, 2019; Eubanks, 2018) interact with datasets and algorithms. It would be irresponsible to accept and trust reference classes generated by a ML system *as is* without further assessment. This assessment necessarily requires auxiliary theories, for example of justice.

Lastly, and fourth, models in the social sciences can change the basic coordinates they describe (Blakeley, 2020). An example is the way "the economy" is measured with *prima facie* neutral indicators such as gross domestic product, the unemployment rate, or the Dow Jones Index, while other indicators—such as the humanity of labor, the impact of economic activities on the environment, or extreme inequalities—are not considered. Taking these as relevant indicators is a choice motivated by political and moral views.

To exemplify this, consider the body mass index (BMI), which is a measure of body fat based on height and weight that categorizes a person along a continuum from underweight to obese. The higher the BMI, the stronger the risk of suffering from heart failure (Khan, 2018). Arguably, the very existence of this model reflects scientific, cultural, political, social, aesthetic, and even religious values and perspectives present in society. Even when the data are not statistically biased *per se*, it's important to be aware that neither the index nor the data are atheoretical in the strict sense, as they determine what counts as an important or promising indicator for detecting risk. To illustrate the difference, consider an example related to the issue of cardiovascular risk. The "social determinants of health" perspective—unlike the BMI—pays primary attention to systemic and structural parameters, such as access to good transportation, education, and housing, which can also be positively or negatively linked to heart disease and stroke (see e.g., WHO, 2010).

5 CONCLUSION

I have explored a series of problems with the "end of theory" view. Prior theory, subjectivity, and values blight the naturalistic effort. The necessary labelling required for training data in supervised ML systems introduces an element of circularity that is unacceptable from a naturalistic point of view. At the same time, assessing the appropriateness of a reference class determined by unsupervised machine learning and profiling techniques requires prior theoretical conceptions of health. ML systems are prone to suffering from sampling and structural biases. These biases are often the result of prior theories and values, which are expressed in the data itself. Previous theories and values are also necessary to recognize and mitigate these problems.

What's more, we do not simply expect an ML system to generate reference classes, which is a computationally trivial task of finding correlations between different variables. Rather, what we expect are *clinically relevant* correlations that enable the system to generate *adequate* reference classes that are also fair and equitable. If we wish to accept a reference class as adequate, we should deem it insufficient to just establish a positive correlation between two or more variables *per se*. We should require explanatory justifications in terms of how and why the system defined a particular reference class (Casacuberta et al., 2022). Yet, unsupervised machine learning systems are said to operate as a "black box" (Holm, 2019), which makes it difficult to comprehend how and based on what reasons the algorithm generates an output.[7]

[7] The link between concepts, explainability, and explanatory justifications merits a richer discussion, but alas, due to space limitations I cannot discuss this matter in further detail. I refer the interested reader to Casacuberta et al., 2022, where my associates and I engage with these themes in depth.

The thrust of medical deliberations is about when to attribute a particular evaluative concept (i.e., healthy or diseased) to a biological state. Take the case of osteoporosis. While its diagnosis largely depends on a quantitative assessment of bone mineral density, the clinical significance of osteoporosis lies in the fractures that arise. The causes of these fractures are multifactorial. To assess the risk of fracture there is a myriad of methods, with different input variables and models that generate different risk estimations (Kanis et al., 2017).

Different conceptions of health enable individuals (medical professionals, patients, citizens in general, and so on) and collectives (such as governments, international and local organizations, patients associations, and so on) to offer reasons in favor or against calling a state or condition "healthy" or "diseased." These critical deliberations have profound implications. The most obvious one is their influence on the contents of classificatory standards, such as the International Statistical Classification of Diseases (ICD). As an illustration, consider the fact that, from the second millennium BC onwards, hysteria was considered a diagnosable physical disease affecting women especially. In 1980, hysterical neurosis was deleted from the DSM, the standard classification of mental disorders (Tasca et al., 2012). No medical professional uses the term "hysterical" anymore. Hysteria is more a reflection of Victorian gender dynamics and oppressive attitudes toward women than anything else. Yet the effects of hysteria once having been an *official* female disease linger on and are suffered by women all over the world on a daily basis.

It is because of this that these critical engagements also fuel the emancipatory collective struggles that seek to remove diagnoses from the official classificatory manuals like ICD and DSM. Besides hysteria, another example concerns the diagnoses that once defined widely prevalent aspects of human sexuality, such as homosexuality, as a mental disorder (Drescher, 2010). Both the classification as a disease as well as the resistance against it reflect scientific perspectives and changing societal views. There is no end of theory.

Defining fraught and value-laden notions such as health is an ongoing project where the discussion is not only about what *is* but also about what *should be*. This dialogical engagement is a question of ethics and politics, not one of finding positive correlations between data; it is not a question of mathematics.

That machine learning won't save naturalism about health from its internal conflicts does not mean that all normatively engaged conceptions of health are equally coherent or comprehensive. Nor does it entail that the search for objectivity must be abandoned—this claim naturally deserves further elaboration, but alas, I lack the space to do so. Suffice it to say that objectivity can still be obtained by evaluating which of any number of "competing theories is more fruitful, better at resolving certain dilemmas, or more able to subject its rival to an effective immanent critique" (Blakeley, 2019).

To end, I wish to present some broader reflections and a call to action. While a fully blown injunction against the use of ML systems for epistemic purposes seems unwarranted, we ought to avoid giving these systems the final word in determining value-laden notions such as health, privacy, gender, trustworthiness, criminality, education, and so on. This is not least because this task requires genuine *judgment*—understood as "deliberative thought, ethical commitment and responsible action"—something no current AI system is capable of (Cantwell Smith, 2019: XV, p. 82). Neither should we—for the sake of neutrality, science or efficiency—abdicate the competence to determine these meanings to the creators and deployers of AI systems rather than society at large. To do so would be to deny the public the possibility of participation, yet public reasoning and discussion are the key to the digital future. Safeguarding the agentic ability to interpret and evaluate the world is a way of retaining fundamental epistemic agency. In other words, we must preserve the public's power to make judgements about what these fraught, normative notions mean. But there's more: If epistemic agency is to be safeguarded, what must also be preserved is the human capacity to discuss what "normality" looks like, that is, discussing what inherently contestable and time-bound reference classes should be the basis for making evaluations related to those normative notions.

6 ACKNOWLEDGEMENTS

The author wishes to thank Sara Pedraz and David Casacuberta, who provided comments on a draft version of this paper. The presentation of this work at the Weizenbaum Conference has been partially funded by the BBVA Foundation through the research project for SARS-CoV-2 and COVID-19 Research in Humanities (Detección y eliminación de sesgos en algoritmos de triaje y localización para la COVID-19).

7 REFERENCES

1. Adamson, A. S., & Smith, A. (2018). Machine learning and health care disparities in dermatology. *JAMA Dermatology*, *154*(11), 1247-1248.
2. Anderson, C. (2008). *The end of theory: The data deluge makes the scientific method obsolete*. Wired Magazine. Retrieved 15/01/22 from https://www.wired.com/2008/06/pb-theory/
3. Benjamin, R. (2019). Race after technology: Abolitionist tools for the new Jim Code. Cambridge: Polity.
4. Blakeley, J. (2020). We built reality: How social science infiltrated culture, politics, and power. Oxford: Oxford University Press.
5. Blakeley, J. (2016). Alasdair MacIntyre, Charles Taylor, and the Demise of Naturalism. Notre Dame: University of Notre Dame Press.
6. Boorse, C. (1977). Health as a theoretical concept. *Philosophy of Science*, *44*(4), 542-573.
7. Boorse, C. (2014). A second rebuttal on health. *Journal of Medicine and Philosophy*, *39*, 683–724.
8. Bowker, G. C., and Star, S. L. (2000). *Sorting things out: Classification and its consequences*. Cambridge: MIT Press.
9. Butler, J. (1990). *Gender Trouble*. London: Routledge.
10. Casacuberta, D., & Vallverdú, J. (2014). E-science and the data deluge. *Philosophical Psychology*, *27*(1), 126-140.
11. Casacuberta, D., Guersenzvaig, A., & Moyano-Fernández, C. (2022). Justificatory explanations in machine learning: for increased transparency through documenting how key concepts drive and underpin design and engineering decisions. *AI & Society*.
12. Cantwell Smith, B. (2019). *The Promise of Artificial Intelligence*. Cambridge: MIT Press.
13. Drescher, J. (2010). Queer diagnoses: parallels and contrasts in the history of homosexuality, gender variance, and the diagnostic and statistical manual. *Arch Sex Behav.* 39(2):427–60.
14. Esteva, A., Kuprel, B., Novoa, R. A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115-118.
15. Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. New York: St. Martin's Press.
16. Falk, D. (2019). How Artificial Intelligence Is Changing Science. Quanta Magazine. Retrieved 15/01/22 from https://www.quantamagazine.org/how-artificial-intelligence-is-changing-science-20190311/
17. Fink, C., Blum, A., Buhl, T., et al. (2020). Diagnostic performance of a deep learning convolutional neural network in the differentiation of combined naevi and melanomas. *Journal of the European Academy of Dermatology and Venereology*, *34*(6), 1355-1361.
18. Heaven, W. D. (2020). DeepMind's protein-folding AI has solved a 50-year-old grand challenge of biology. *MIT Technology Review*. Retrieved 10/01/2022 from https://www.technologyreview.com/2020/11/30/1012712/deepmind-protein-folding-ai-solved-biology-science-drugs-disease/
19. Holm, E.A. (2019). In defense of the black box. *Science*, *364*(6435), 26-27.

20. Huntington, A., & Gilmour, J. A. (2005). A life shaped by pain: Women and endometriosis. *Journal of Clinical Nursing, 14*(9), 1124-1132.

21. Kanis, J. A., Harvey, N. C., Johansson, H., et al. (2017). Overview of Fracture Prediction Tools. *Journal of clinical densitometry: the official journal of the International Society for Clinical Densitometry*, *20*(3), 444–450.

22. Khan, S. S., Ning, H., Wilkins, J. T., et al. (2018). Association of Body Mass Index With Lifetime Risk of Cardiovascular Disease and Compression of Morbidity. *JAMA cardiology*, *3*(4), 280–287.

23. Kingma, E. (2014). Naturalism about health and disease: Adding nuance for progress. *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine, 39*(6), 590-608.

24. Male, V. (2021). Menstrual changes after COVID-19 vaccination *BMJ*, 374:n2211.

25. Ridker, P.M., Cook, N.R., Lee, et al. (2005). A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. *New England Journal of Medicine. 352*(13):1293–1304.

26. Schawinski, K., Turp, D.M., & Zhang, C. (2018). Exploring galaxy evolution with generative models. *Astronomy & Astrophysics*, 616, L4.

27. Tasca, C., Rapetti, M., Carta, M. G., & Fadda, B. (2012). Women and hysteria in the history of mental health. *Clinical practice and epidemiology in mental health*, 8, 110–119.

28. USGAO – United States General Accounting Office. (2001). *Drug Safety: Most Drugs withdrawn in Recent Years had Greater Health Risks for Women*. Washington, DC: US Government Publishing Office.

29. WHO - World Health Organization. (2010). *A conceptual framework for action on the social determinants of health.* World Health Organization. Retrieved 15/01/22 from https://apps.who.int/iris/handle/10665/44489

Proceedings of the Weizenbaum Conference 2022:
Practicing Sovereignty. Interventions for Open Digital Futures

REUSE SOFTWARE

MAKING COPYRIGHT AND LICENSING COMPLIANCE EASIER FOR EVERYONE

Lasota, Lucas
Humboldt University of Berlin
Free Software Foundation Europe
Berlin, Germany
lucas.lasota@hu-berlin.de

KEYWORDS

software license; free and open source software; copyright; compliance

DOI: 10.34669/wi.cp/4.6

ABSTRACT

Best practices for displaying data and metadata pertaining to software licensing and copyright are currently unharmonized. The multiple competing licensing requirements for communicating the chosen license of a software project and its copyright holders increase the compliance burden on project maintainers, especially for smaller free and open source (FOSS) ones. The “REUSE Software” initiative aims to remediate this situation by defining a set of easy-to-implement best practices for declaring copyright and licensing in an unambiguous, human- and machine-readable way, so that the information is preserved when the file is copied and reused by third parties. REUSE specifications facilitate management policies for digital commons, improving data and metadata communication for individuals, communities, governments, and businesses.

1 INTRODUCTION[8]

Digital transformation necessarily involves copyright and licenses, because software, the backbone element of digital technologies, is regulated by copyright.[9] The re-usability of software should be authorized by licenses or statutory copyright limitations or exceptions.[10] The multiple competing requirements for communicating the chosen license and the copyright holders increase the compliance burden on project maintainers, especially for smaller free and open source software (FOSS)[11] ones. The "REUSE Software" initiative[12] defines best practices for declaring copyright and licensing in an unambiguous, human- and machine-readable way, so that the information is preserved when the file is copied and reused by third parties. REUSE specifications aim to facilitate and improve management policies for the digital commons, improving data and metadata communication for individuals, communities, governments, and businesses.

2 CHALLENGES FOR COMPLIANCE

The more external components a software code encompasses, the harder it is to keep an overview of the copyright holders and their licensing choices. Since FOSS licenses are public documents that are shared openly, often by millions of users worldwide, their implementation generally does not involve negotiation among the parties. Therefore, proper information regarding the governing license is crucial to avoid legal (Synopsys, 2019) and security risks (Haddad, 2018). This is especially problematic for FOSS projects, as large public code repositories mean a decreased number of licensed repositories (Balter, 2015). Moreover, license proliferation fragments the requirements for copyright and license notices (OSI, 2006). Software projects incorporating content elements—as text, images, and videos—face an additional layer of complexity with content licensing compliance.[13]

How copyright and license information should be displayed depends on copyright law and license requirements. Especially important are notices for reciprocal licenses (also known as copyleft), as they require the derivative work to be licensed under the same licensing terms, which directly impacts license compatibility (Ku Wei Bin, Lasota & Jaeger, 2022). Although FOSS licenses

[8] This paper does not necessarily reflect the views of any organization the author may represent. The author thanks Richard Schmeidler and the reviewers for the proofreading and comments on the text.
[9] For the European Union, see Art. 1(1) of Software Directive (2009/24/EC) from 23 April of 2009.
[10] Regarding statutory exceptions, see Blázquez, Cappello & Valais, 2017.
[11] The definitions of free and open source software are taken respectively from the Free Software Foundation and Open Source Initiative. See Ku Wei Bin, Lasota & Jaeger, 2022, p. 10.
[12] See the project's web portal. Available at: https://reuse.software/ Retrieved on 30.06.22.
[13] See, for instance, the Creative Commons recommendations for applying a license to creative works. Available at: https://wiki.creativecommons.org/wiki/Marking_your_work_with_a_CC_license Retrieved on 30.06.22.

in general provide information on how the license notices should be applied, the vastly diverse recommendations remain unharmonized.

3 REUSE: SETTING HARMONIZED BEST PRACTICES

The REUSE best practices enable humans and machines alike to add and read data and metadata regarding licenses and copyright notices. They intend to relieve the license compliance burden for software projects and improve standardization for data and metadata transfer. This is relevant for:

- Individual developers, because it provides them a precise and easy-to-implement way to apply correct terms of license and copyright notices.
- Digital communities, because it improves how data and metadata for software re-usability is communicated.
- Academia, because it improves the re-usability of software in a safe and clear way in research projects.
- The public sector, because it fosters best practices for dealing with license and copyright notices, improves interoperability among agencies, and encourages open government.
- Commercial entities, because it allows them to optimize their software bill of materials and simplify development workflow.

4 REUSE SPECIFICATIONS

REUSE's core specifications are based on SPDX,[14] an open standard for communicating software bill of material information, including components, copyrights, licenses and security references. SPDX maintains a license list,[15] which defines standardized identifiers for a wide spectrum of commonly found licenses and exceptions used in FOSS, data, hardware, or documentation. SPDX was designed to provide "a common language and vocabulary to express security, licensing, and copyright information for products, components, packages, files and code snippets, enable tools to be created and facilitate the introduction of compliance automation." (Haddad, 2018, p. 154) The combination of these standards enables the REUSE's three-step compliance procedure:

[14] Software Package Data Exchange (SPDX) is an international ISO open standard (ISO/IEC 5962:2021) managed by the Linux Foundation.
[15] The SPDX License List includes a standardized short identifier, the full name, the license text, and a canonical permanent URL for each license and exception. Available at: https://spdx.org/licenses, retrieved on 30.06.22.

- Choosing and applying a license by downloading from the SPDX list the license text and information. This data should be stored in a *LICENSES* directory in the source code repository.
- *Providing to every single file a copyright and license header based on the SPDX standard:*
  - *# SPDX-FileCopyrightText: [year] [copyright holder] <[email address]>*
  - *# SPDX-License-Identifier: [identifier]*
- Confirming compliance with the REUSE toolkit,[16] which also is capable of automating the two previous steps.

5 RESOURCES AND SUPPORTERS

Software license compliance is a complex and vast area populated with a multitude of initiatives and tools to help compliance efforts. REUSE, which has a community-based approach, collaborates with several complementary projects[17], such as ClearlyDefined[18], OpenChain[19] and FOSSology[20]. In addition, for developers, there is a series of resources for easy engagement and adoptions, an open mailing list for discussion and deliberation, extensive FAQs, and a constantly updated toolkit with compliance tools, API checks, and provision for numerous CI/CD solutions.

Although it is not possible to know exactly the number of adopters, by February 2023, 1443 software repositories using the REUSE API are successfully implementing and following the best practices. REUSE had been adopted by the Linux Kernel, and several large companies. The specifications are also a central element in the compliance workflow for the European Commission's Next Generation Internet Initiative,[21] serving as consortium best practices for software and research projects developing human-centric technologies for the future of the Internet.

---

[16] See the tool's dedicated section on the REUSE website. Available at: https://reuse.software/faq/#tool, retrieved on 30.06.222.
[17] For an overview of complementary initiatives, see: https://reuse.software/comparison/, retrieved on 30.06.22.
[18] ClearlyDefined is an Open Source Initiative incubator project. The goals of the project are to collect and display meta and security information about a large number of software and data projects distributed on different package registries. It also motivates developers and curators to extend data about a project's licensing and copyright situation. REUSE in comparison concentrates on fixing the problem at the file level for individual projects. See: https://clearlydefined.io/about. Retrieved on 07.02.23.
[19] The OpenChain Project is focused on building trust in the free software supply chain. OpenChain focuses on making free software license compliance more transparent, predictable, and understandable for participants in the software supply chain. OpenChain recommends REUSE as one component to increase clarity of the licensing and copyright situation, but has higher requirements to achieve full conformance. See: https://www.openchainproject.org/, retrieved on 07.02.23.
[20] FOSSology is a toolkit for Free Software compliance, stores information in a database, and includes license, copyright and export scanners. It is more complex than REUSE and its helper tool and rather optimized for compliance officers and lawyers. REUSE instead intends to have all licensing and copyright information stored in or next to the source files to safeguard this information when reused elsewhere. See: https://www.fossology.org/about/, retrieved on 07.02.23.
[21] For a detailed overview of the initiative, see the NGI0 Zero website. Available at: https://www.ngi.eu/ngi-projects/ngi-zero/, retrieved on 30.06.22.

6 REFERENCES

1. Balter, Ben (2015). *Open source license usage on GitHub.com*. GitHub Blog. Retrieved on 26.06.2022. Available at: https://github.blog/2015-03-09-open-source-license-usage-on-github-com/
2. Blázquez, Cappello & Valais (2017). *Exceptions and limitations to copyright*. IRIS Plus, European Audiovisual Observatory, Strasbourg.
3. Haddad, Ibrahim (2018). *Open Source Compliance in the Enterprise*. 2nd ed. The Linux Foundation: San Francisco.
4. Ku Wei Bin, G., Lasota, L. and Jaeger, T. (2022). *Free and Open Source Software Licensing: Frequently Asked Questions - Next Generation Internet Legal To-Dos*. Berlin: Free Software Foundation Europe.
5. Open Source Initiative (2006). *Report of License Proliferation Committee and draft FAQ*. OSI Website. Retrieved on 26.06.2022. Available at: https://opensource.org/proliferation-report
6. Synopsys (2019). *Top open source licenses and legal risk for developers*. Retrieved on 30.06.22. Available at: https://www.synopsys.com/blogs/software-security/top-open-source-licenses/

Proceedings of the Weizenbaum Conference 2022:
Practicing Sovereignty. Interventions for Open Digital Futures

DIGITAL INCLUSION OF LOW-LITERATE ADULTS

CHALLENGING THE SEQUENTIAL UNDERPINNINGS OF THE DIGITAL DIVIDE

Smit, Alexander
University of Groningen
Groningen, the Netherlands
a.p.smit@rug.nl

Swart, Joëlle
University of Groningen
Groningen, the Netherlands
j.a.c.swart@rug.nl

Broersma, Marcel
University of Groningen
Groningen, the Netherlands
m.j.broersma@rug.nl

KEYWORDS

digital inclusion; digital divide; participation; digital literacies; low-literacy

DOI: 10.34669/wi.cp/4.7

ABSTRACT

Contemporary models of digital inclusion and the digital divide assume that developing the digital literacy that enables individuals to participate in society is a sequential and linear process that is more or less similar for all individuals in all contexts and requires basic linguistic skills. This paper challenges these understandings, arguing that such a technical, normative perspective excludes marginalized and disadvantaged publics, such as low-(digital) literate citizens. Based on a longitudinal ethnographic study of low-literate Dutch adults, we show that the often-described causal relation between (digital) literacies, (digital) participation, and (digital) inclusion is not as evident as it seems and neglects the important socio-cultural contexts through which (digital) literacies are often gained and enacted in everyday practice. Consequently, we argue that current conceptualizations of (digital) inclusion and (digital) participation need to be rethought in terms of the limitations, potential, and capabilities of low-literate people.

1 INTRODUCTION

As more and more aspects of society become digitized, citizens are increasingly expected to participate digitally, a process that increases digital inequalities. It is often thought that (digital) literacies facilitate participation and that, hence, (digital) literacies should be understood as the gateway towards fostering (digital) inclusion (Hargittai and Hinnant, 2008; van Dijk, 2020). However, the relation between digital (low-)literacy and digital inclusion is complex and remains understudied, especially in light of marginalized publics (Ragnedda, 2016; Selwyn, 2004). Most current models of the digital divide employ a rather linear, sequential, and instrumental rationale regarding the use of digital technologies (Van Deursen, Helsper, and Eynon, 2016; van Dijk, 2020). Van Deursen et al. (2016), for example, formulated the concept of *sequential digital exclusion*, thereby distinguishing between several sequential levels of inequality where a lack of digital literacies prevents digital participation (e.g., access, skills, usage, motivation, etcetera). Such studies understand digital inclusion in a technical sense and presuppose that basic linguistic skills are necessary to participate digitally. Yet, this is problematic, as it implies a normative understanding of (digital) participation, where an individual is only able to participate if they possess the necessary basic (digital) literacy skills to make use of digital media. Hence, this understanding is intrinsically exclusionary for marginalized publics, such as low-literate adults.

This group's use of information and communications technologies (ICTs) largely differs from that of more generic publics that have higher levels of traditional and digital literacy (Grotlüschen et al., 2019; Tsatsou, 2021). For example, low-literate adults may rely upon third-party actors such as literacy supporters (Grotlüschen et al., 2019), or digital care workers (Kaun and Forsman, 2022) to use ICTs. Additionally, current understandings of (digital) participation and the digital divide largely neglect the situated socio-cultural contexts through which disadvantaged publics participate—for example, how they make use of digital media in more affective and social ways (Buddeberg, 2016; Yilmaz, 2016). As such, dominant understandings of digital inclusion/exclusion and the digital divide must be scrutinized and reconsidered in light of a more inclusive conception that focuses on the potential capabilities of low-literate adults, and the limitations they are confronted with in their everyday lives. Hence, this article problematizes current understandings of participation, digital inclusion, and the digital divide in light of low-literate Dutch adults, a subgroup that is heavily understudied and runs the risk of falling behind in an ever-increasing digitalization of society.

Building upon an ethnographic study consisting of participant observations and in-depth interviews with low-literate Dutch adults (N=73), this paper presents an analysis of the sequential and hierarchical underpinnings of contemporary models of digital inclusion and the digital divide. We

challenge three assumptions that underly current understandings of digital literacies and the digital divide. First, we argue that sequential models of the digital divide are problematic when applied to the situated (digital) practices of low-literate Dutch adults. Second, current understandings of literacy mostly draw upon autonomous models of literacy education (Street, 2003), which ignores the social contexts through which (digital) literacies are often gained and enacted. We argue that this becomes problematic when translated to the everyday practices of low-literate individuals, as studies show that such adults often make use of their social network to be able to participate in society at large (Grotlüschen et al., 2019; Kaun and Forsman, 2022). However, such social actors are not considered in models of the digital divide and/or digital inclusion (van Deursen et al., 2014; van Dijk, 2020). Third, we scrutinize the prioritization of digital literacies over traditional literacy and the neglect of socio-cultural contexts when learning about and enacting digital literacies in practice.

Drawing from these three points, this article shows how and why current sequential understandings misinterpret the participatory practices of low-literate adults, which entail various digital literacies and (linguistic) limitations that low-literate adults experience in their daily lives. We argue that the often-described causal relation between (digital) literacies and (digital) inclusion is not as evident as it seems and needs to be rethought in terms of the limitations, potential, and capabilities of low-literate citizens.

2 CHALLENGING THREE UNDERPINNINGS OF THE DIGITAL DIVIDE

1.1 SEQUENTIALITY OF THE DIGITAL DIVIDE

Current work on digital inequalities identifies three digital divides at distinct levels, relating to: (1) the access and frequency of use, (2) successive kinds of access, users' skills and diversity of use, and finally, (3) the benefits and potential outcomes of media use (Helsper, 2012; van Dijk, 2020). The first-level digital divide refers to inequalities in access to digital technologies related to different (economic) backgrounds, often divided in a binary manner of haves and have nots (e.g., Hargittai, 1999). The second-level digital divide relates to differences in digital skills and diversity of use in regard of four successive kinds of access: (1) motivational, (2) material, (3) skills, and (4) usage access (Van Dijk 2005, 2020). The third-level digital gap shifts the focus to the different outcomes achieved after using digital technologies (e.g., Helsper 2012; Van Deursen et al. 2017). Additionally, current (quantitative) research primarily offers selective insights into these three levels of the digital divide, and often engages with more 'visible' parts of the populace (Goedhart, Verdenk & Dedding, 2022; Helsper & Reisdorf, 2017; van Deursen & van Dijk, 2015). Most contemporary studies focus on the second and/or third level and/or the transition from the second towards the third level, because

in most instances a divide in the first level of access is closed when someone acquires a digital device (Hargittai, Piper and Morris, 2018; Helsper 2012; Scheerder, van Deursen, van Dijk, 2019). Thus, it is often argued that when the first level is addressed, users can move on to the second level and then to the third level. For example, when a citizen buys a laptop or a smartphone, this allows him/her to develop digital skills by trial-and-error, which then may lead to being able to gain tangible outcomes from such digital tasks (see Figure 1). This is a somewhat oversimplified illustration of how the digital divide is conceptualized. However, policy and education often presuppose this sequential model of the digital divide in the context of digital inequality (Mariën et al., 2016). Additionally, reports on the digital divide are also often constructed on similar normative underpinnings that digital media usage will—almost naturally—result in positive development, engrained in a techno-solutionist narrative (Helsper, 2012; 2021).

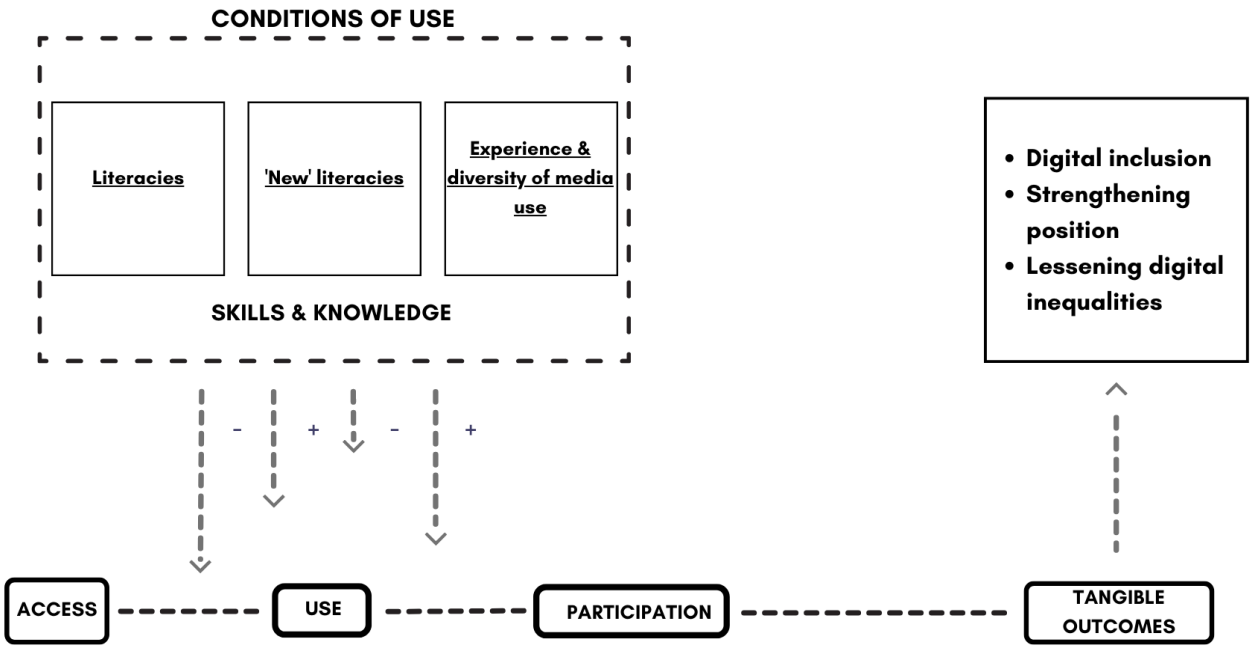


Figure 1: Sequential model of the digital divide

This sequential way of thinking about the digital divide implies a hierarchical order of exclusion, where citizens are excluded in the order of the three levels of the digital divide (Helsper, 2012). Furthermore, this way of thinking about the digital divide is often understood in terms of the transition from one level to another and not in relation to differences within the levels themselves. The latter would be relevant, for example, when comparing someone acquiring a smartphone and learning to use it with someone acquiring a laptop or PC, which has entirely different hardware and software and prebuilt norms. A sequential understanding is implicitly incorporated within this level-to-level progression: Poor technical skills, for example, mean that an individual will not even have the

opportunity to perform other informational skills and tasks (Friemel et al., 2021). Finally, these models assume that this hierarchy is largely equal for all individuals across contexts, while people's situated position in interrelated structures of power is often negated (Zheng & Walshman, 2021). For example, when members of the general public gain access to digital media and want to develop digital skills—i.e., when buying a laptop or smartphone—they may read the manual accompanying the device if something is unclear or search the internet for additional guidelines on how to make use of the device and educate themselves.

However, when this process is applied to the context of a low-literate user, with the associated limitations and capabilities, this becomes far more complex, as that individual cannot simply educate themselves because of literacy limitations. According to the sequential model, this means that there is no way for these low-literate individuals to advance to the second and third level of the digital divide. Yet, our observations show that even when they lack access, low-literate adults can develop skills or achieve outcomes not deemed possible by current models of the digital divide (Van Dijk, 2020). For example, when low-literate individuals make use of *literacy supporters* (Grotlüschen et al., 2019) or *digital care workers* (Kaun and Forsman, 2022), they do not have direct access to ICTs; however, they are still able to participate with the help of third parties. Thus, the process of developing (digital) literacies that facilitate digital inclusion is not necessarily hierarchical or linear; the order between these levels may differ depending on users' personal, technological, socio-cultural, economic, and political contexts. As such, the sequence of action is not primarily hierarchical or linear; rather, it is recursive and fluid.

To our knowledge, only a few studies have taken this nonsequential relationship between the dimensions of digital inequalities into account and have explored how it relates to disadvantaged publics (Buddeberg, 2016; Friemel et al. 2021; Kaun and Forsman, 2022; Tirado-Morueta et al. 2017; Wei et al. 2011). One reason might be because some of the basic inequalities (e.g., concerning access) have been considered solved among most publics. However, they are still very relevant within the context of low-literate adults. Such publics typically do not possess the same economic resources to attain media devices as the more general publics and do not have the basic literacy skills to gain knowledge from and through such digital devices to make effective use of them. As such, more work is needed focusing on this first level, while simultaneously relating it to the second and third levels within personal, technological, socio-cultural, economic, and political dimensions.

1.2 FORGETTING THE SOCIAL CONTEXT

Digital literacies are understood as an important aspect of being able to participate digitally; however, the social context through which digital literacies are learned, practiced, and appropriated should

not be forgotten (Ananiadou & Claro 2009). Prior work on digital inequalities assumes that digital literacies are neutral and technical concepts that can be used by drawing upon a monolithic set of skills (Gee, 1991). This goes back to what Brian Street calls the "autonomous" model of literacy education (Street, 1985). Street posits that two dominant forms of literacy exist: the "autonomous model" and the "ideological model." He hence distinguishes between literacy events and practices (Street, 1985, p. 77). He describes the autonomous model as:

> Introducing literacy to poor, "illiterate" people, villages, urban youth etc. will have the effect of enhancing their cognitive skills, improving their economic prospects, making them better citizens, regardless of the social and economic conditions that accounted for their "illiteracy" in the first place. I refer to this as an "autonomous" model of literacy. The model, I suggest, disguises the cultural and ideological assumptions that underpin it so that it can then be presented as though they are neutral and universal and that literacy as such will have these benign effects (Street, 2003, p. 77).

This autonomous approach thus imposes western conceptions of literacy on to other cultures, or within countries with different socio-economic classes (Street, 2003). This perspective has been challenged in recent decades, as studies showed that in practice literacy has different effects depending on the context through which it is enacted (Gee, 1991) and differs depending on the socio-cultural arrangements it draws upon (Street, 2003).

Scholars have therefore increasingly adopted the ideological model of literacy, which constructs a more situated and "culturally sensitive view of literacy practices as they vary from one context to another" (Street, 2003, p. 78). As such, this understanding relates more to the livelihoods and societal positions of low-literate citizens, as these marginalized groups face very different issues regarding participation in society depending on differences in their ethnicity, age, disability, gender, educational level, and socio-economic position. The ideological model draws on the understanding that literacy is first and foremost a social practice and not a purely neutral and/or technical tool to understand and make use of (digital) texts. As Street notes: "the ways in which people address reading and writing are themselves rooted in conceptions of knowledge, identity, and being (Street, 2003, p. 78). This underpins the contextual and situational nature of literacies, which is often neglected when we explore how such traditional literacies translate into digital ones that foreground a solely technical and neutral understanding of literacy.

While digital literacies indeed help to make use of digital media, this undermines all of the other dimensions that collectively shape how users of digital media understand and enact the digital world (Friemel et al. 2021). For example, while autonomous models of literacy put cognitive skills at their core and presuppose that reading and writing abilities are necessary for digital participation, this

does not take into account the personal, technological, socio-cultural, economic, and political contexts through which they are translated into everyday (digital) practices. Additionally, actors providing social support to low-literate individuals in navigating the digital society are very important in this process, as they act as third parties who take them by the hand and show them how digital infrastructures work (Grotlüschen et al., 2019; Kaun and Forsman, 2022). Such third-party actors are largely forgotten in current understandings of participation, digital inclusion, and the enactment of digital literacies, but they are highly influential in how ICTs are used in everyday life (Grotlüschen et al., 2019; Kaun and Forsman, 2022).

In addition, our research shows that the affective dimension of human-technology relations is very influential in the context of low-literate adults, as they fill their gaps in cognitive skills regarding the digital with more tacit modes of knowing, that is, gut-feeling, fear, doubt, etcetera. This is another factor largely neglected in current models of the digital divide, yet it seems to be of great importance when talking with low-literate adults about their usage of ICTs. In this sense, the ideological model is more applicable to the multi-dimensionality of digital literacies in the context of its publics and how they enact digital literacies in situated daily practices, originating from their socio-economic conditions. This gives a more contextualized understanding of why and how digital (il)literacy affects inequality and what role (digital) literacies play in diminishing such inequalities in an increasingly digital society where more and more citizens run the risk of being left behind.

1.3 THE NEED FOR TRADITIONAL LITERACY IN THE DEVELOPMENT OF DIGITAL LITERACIES

Dominant understandings regarding digital literacies presuppose that basic literacy is a necessary starting point for the development of digital literacies (Friemel et al., 2021; van Dijk, 2020). Most low-literate publics we conversed with acknowledge that possessing basic linguistic skills is a prerequisite for participating as democratic citizens. However, they also note that the more digital literacies they gain, the more they can circumvent their traditional issues with linguistic proficiency and leverage affordances of media and software to enlarge their capabilities with these technologies and strengthen their societal position (Smit, Swart, and Broersma, forthcoming). Thus, the development of digital literacies does not always follow a linear path towards digital participation, digital inclusion and so forth, and can also potentially be leveraged for digital and/or societal non-participation resulting in (digital) exclusion. For example, as we found in our study, migrants and/or refugees may use Google Translate to speak in their native language and let their smartphone translate their native language to another language, bypassing the need to learn the language of the country they currently reside in (Smit, Swart and Broersma, forthcoming). In this way, they consciously exclude themselves from broader society and its cultural codes (language).

This finding aligns with results from studies regarding differences in internet use by diverse publics (Scheerder, van Deursen, van Dijk, 2017; Scheerder, van Deursen, van Dijk, 2019; Van Dijk, 2020), which show that people from different social classes, of different ages, with different genders, and from different ethnic, cultural, and other backgrounds are increasingly using the internet differently. The structural divide observed here is called the usage gap: People with high education levels and social class status use more informational, educational, work, and career enhancing applications, while people with low education levels and social class primarily use apps that offer entertainment, chat or simple communication, and e-shopping (Hargittai and Shafer, 2006; Van Dijk, 2020; Yilmaz, 2016). This partially stems from a difference in skillset, mindset, and affective attitude towards media in general. Affective attitude towards digital media is especially important for low-literate adults, as our results show that low-literate adults prioritize emotions, such as gut-feeling, intuition, and fear in who and what to (dis-)trust—for example, when arranging financial matters through e-banking. Hence, the dominant perspective on digital inclusion and participation as an individualized technical endeavor is not in line with the socially situated everyday practices of low-literate adults. This perspective needs to shift from a top-down prescriptive conceptualization of inclusion and participation towards a bottom-up socially situated perspective. Instead of only talking about how marginalized publics should participate and be included, we argue that we should rather pay attention to what these groups themselves deem important in how participation and inclusion manifest in their everyday life.

3 CONCLUSION

Now that economic and social inequalities are rising in large parts of the world, we are confronted with the increasing complexity of closing the digital divide as the digitalization of society progresses. The digital cannot be isolated from the social and vice-versa, so we need to simultaneously fight against digital and social inequality. The socio-economic, cultural, and personal situations of marginalized publics must be centered within policies and pedagogies if we are to simultaneously battle social and digital inequalities in ways that account for what these publics themselves deem important for how to thrive in societies. We should ask whether, for such disadvantaged publics, a digital-by-default society is desirable. It is crucial to explore which skills and attitudes are needed to include marginalized publics into ever-expanding digital societies. In doing so, we can develop better situated and contextualized pedagogies that center the users of digital media from a bottom-up perspective instead of enforcing norms and learning outcomes from a top-down perspective that does not apply to the personal socio-cultural situatedness of these publics. Additionally, more studies are needed to understand the relationships between social and digital inequalities and how they re-enforce

one another. Hence, by centering the potential and possibilities of disadvantaged and/or marginalized publics in how they can participate in situated ways, instead of solely focusing on their limitations, we can develop better suited educational systems, pedagogies, and policies to empower them to participate in ways that are in line with their capabilities and affective dispositions towards digital media.

4 ACKNOWLEDGMENTS

This work was supported by the Dutch Research Council (NWO), Grant nr. 410.19.008.

5 REFERENCES

1. Ananiadou, K., & Claro, M. (2009). 21st century skills and competences for new millennium learners in
2. OECD countries (OECD Education Working Papers, No. 41). Paris, France: OECD Publishing. https://doi.org/10.1787/218525261154
3. Buddeberg, K. (2016). Hauptergebnisse der quantitativen Teilstudie. In W. Riekmann, K. Buddeberg, & A. Grotlüschen (Eds.), Alphabetisierung und Grundbildung: Vol. 12. Das mitwissende Umfeld von Erwachsenen mit geringen Lese- und Schreibkompetenzen. Ergebnisse aus der Umfeldstudie (pp. 61–78). Münster: Waxmann.
4. Calvani, A., Fini, A., Ranieri, M., & Picci, P. (2012). Are young generations in secondary school digitally competent? A study on Italian teenagers. Computers & Education, 58(2), 797–807. https://doi.org/10.1016/j.compedu.2011.10.004
5. Carpentieri, J. D. (2015). Adding new numbers to the literacy narrative: Using PIAAC data to focus on literacy practices. In M. Hamilton, B. Maddox, & C. Addey (Eds.), Literacy as numbers: Researching the politics and practices of international literary assessment (pp. 93–110). Cambridge: Cambridge University Press.
6. van Deursen, A. J. A. M., Helsper, E. J., & Eynon, R. (2014). Measuring digital skills: From digital skills to tangible outcomes project report. London, UK: London School of Economics and Political Science.
7. van Deursen, A. J. A. M., & van Dijk, J. A. G. M. (2015). Internet skill levels increase, but gaps widen: A longitudinal cross-sectional analysis (2010–2013) among the Dutch population. Information, Communication & Society, 18, 782–797. https://doi.org/10.1080/1369118X.2014.994544
8. van Deursen, A. J. A. M., Helsper, E. J., & Eynon, R. (2016). Development and validation of the Internet Skills Scale (ISS). Information, Communication & Society, 19(6), 804–823. https://doi.org/10.1080/1369118X.2015.1078834
9. van Deursen, A. J. A. M., Helsper, E. J., Eynon, R., & van Dijk, J. A. G. M. (2017). The compoundness and sequentiality of digital inequality. International Journal of Communication, 11, 452–473.
10. van Dijk, J. A. G. M. (2005). The deepening divide: Inequality in the Information Society. SAGE Publications.
11. van Dijk, J. A. G. M. (2020). The digital divide. Polity Press.
12. Friemel, T., Frey, T., & Seifert, A. (2021). Multidimensional Digital Inequalities: Theoretical Framework, Empirical Investigation, and Policy Implications of Digital Inequalities among Older Adults. Weizenbaum Journal of the Digital Society, 1(1), w1.1.3. https://doi.org/10.34669/wi.wjds/1.1.3
13. Gee, J. P. (2010). New digital media and learning as an emerging area and "worked examples" as one way forward. The John D. and Catherine T. Macarthur Foundation reports on digital media and learning. Cambridge, Mass.: The MIT Press.
14. Gee, J.P. (1991). Social Linguistics: Ideology in Discourses, Falmer Press: London
15. Goedhart, N.S., Verdonk, P., & Dedding, C. (2022). "Never good enough." A situated understanding of the impact of digitalization on citizens living in a low socioeconomic position. Policy & Internet, 14, 824-844

16. Gorur, R. (2015). Assembling a sociology of numbers. In M. Hamilton, B. Maddox, & C. Addey (Eds.), Literacy as numbers: Researching the politics and practices of international literary assessment (pp. 1–16). Cambridge: Cambridge University Press.
17. Grotlüschen A, Buddeberg K, Redmer A, Ansen H, Dannath J. (2019). Vulnerable Subgroups and Numeracy Practices: How Poverty, Debt, and Unemployment Relate to Everyday Numeracy Practices. Adult Education Quarterly. 2019;69(4):251-270. https://doi.org/10.1177/0741713619841132
18. Hargittai, E., & Shafer, S. (2006). Differences in actual and perceived online skills: The role of gender. Social Science Quarterly, 87(2), 432–448. https://doi.org/10.1111/j.1540-6237.2006.00389
19. Hargittai, E., & Hinnant, A. (2008). Digital Inequality: Differences in Young Adults' Use of the Internet. Communication Research, 35(5), 602–621. https://doi.org/10.1177/0093650208321782
20. Hargittai, E. (1999). Weaving the western web: Explaining differences in Internet connectivity among OECD countries. Telecommunications Policy, 23(10–11), 701–718. https://doi.org/10.1016/S0308-5961(99)00050-6
21. Hargittai E, Piper A.M. and Morris M.R. (2018). From internet access to internet skills: digital inequality among older adults. Universal Access in the Information Society. Epub ahead of print 3 May. https://doi.org/10.1007/s10209-018-0617-5.
22. Helsper, E. J. (2012). A corresponding fields model for the links between social and digital exclusion. Communication Theory, 22(4), 403–426. https://doi.org/10.1111/j.1468-2885.2012.01416.x Communication Theory 22(4): 403–426.
23. Helsper, E. J., & Reisdorf, B. C. (2017). The emergence of a 'digital underclass' in Great Britain and Sweden: Changing reasons for digital exclusion. New Media & Society, 19(8), 1253–1270. https://doi.org/10.1177/1461444816634676
24. Kaun, A., & Forsman, M. (2022). Digital care work at public libraries: Making Digital First possible. New Media & Society, 14614448221104234.
25. Koltay, T. (2011). The media and the literacies: Media literacy, information literacy, digital literacy. Media, Culture & Society, 33(2), 211–221. https://doi.org/10.1177/0163443710393382
26. Lanvin, B., & Passman, P. (2008). Building e-skills for the Information Age, Chapter 1.6 of "The Global Information Technology Report 2007-2008".
27. Mariën, I., Heyman, R., Salemink, K., & van Audenhove, L. (2016). Digital by default: Consequences, casualties and coping strategies. In J. Servaes, & T. Oyedemi (Eds.), Social inequalities, media and communication: Theory and roots. Exington Books.
28. Ragnedda, M. (2016). The Third Digital Divide: A Weberian Approach to Digital Inequalities (1st ed.). Routledge. https://doi.org/10.4324/9781315606002
29. Scheerder A., van Deursen AJAM and van Dijk JAGM. (2017). Determinants of Internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. Telematics and Informatics 34(8): 1607–1624.
30. Scheerder, A. J., van Deursen, A. J., & van Dijk, J. A. (2019). Negative outcomes of Internet use: A qualitative analysis in the homes of families with different educational backgrounds. The information society, 35(5), 286–298.

31. Scheerder, A. J. (2019). Inevitable inequalities? Exploring Differences in Internet Domestication Between Less and Highly Educated Families.
32. Selwyn, N. (2004). Reconsidering Political and Popular Understandings of the Digital Divide. New Media & Society, 6(3), 341–362. https://doi.org/10.1177/1461444804042519
33. Smit, A.P., Swart, J.A.C., Broersma, M.J. Forthcoming; not yet published. Digital Inclusion of Low Literate Dutch Adults: Digital Literacies and Tactics of Media Use
34. Street, B. (1984) Literacy in Theory and Practice Cambridge: CUP
35. Street, B. (1995). Social literacies: Critical approaches to literacy in development, ethnography and education. New York, NY: Longman.
36. Street, B. (1997). The implications of the "new literacy studies" for literacy education. English in Education, 31(3), 45–59.
37. Street, B. (2003). Current Issues in Comparative Education, Teachers College, Columbia University Current Issues in Comparative Education, Vol. 5(2)
38. Tsatsou, P. (2021). Vulnerable people's digital inclusion: intersectionality patterns and associated lessons. Information, Communication & Society, 1–20.
39. Tirado-Morueta, R., Mendoza-Zambrano, D. M., Aguaded-Gómez, J. I., & Marín-Gutiérrez, I. (2017). Empirical study of a sequence of access to Internet use in Ecuador. Telematics and Informatics, 34(4), 171–183.
40. Yilmaz, F. G. K. (2016). The relationship between metacognitive awareness and online information searching strategies. Pegem Journal of Education & Instruction, 6(4), 447–468. https://doi.org/10.14527/pegegog.2016.022
41. Wei K.K., Teo H.H., Chan H.C., et al.. (2011). Conceptualizing and testing a social cognitive model of the digital divide. Information Systems Research 22(1): 170–187.
42. Zheng, Y., & Walsham, G. (2021). Inequality of what? An intersectional approach to digital inequality under Covid-19. Information and Organization, 31(1), 100341.

Proceedings of the Weizenbaum Conference 2022:
Practicing Sovereignty. Interventions for Open Digital Futures

COMMUNITY-GOVERNED AND COMMUNITY-PAID PUBLISHING

RESILIENT SUPPORT FOR INDEPENDENT OPEN ACCESS JOURNALS

Wrzesinski, Marcel
Alexander von Humboldt Institute for Internet
and Society
Berlin, Germany
marcel.wrzesinski@hiig.de

KEYWORDS

open access; journals; publishing; resilience

DOI: 10.34669/wi.cp/4.8

ABSTRACT

Community-driven open access journals foster the idea of a biblio-diverse publishing ecosystem and challenge the prevalent commercialization of academic publishing. But despite their importance, their existence is threatened. With little to no budget they operate mostly on "gifted labor" (Adema/Moore, 2018, 8) by their editorial teams and free support by public infrastructures. The first part of this article describes the model, key functions, and governance principles of community-driven open access journals within the business of global academic publishing. In promoting fair, resilient, and gratis open access, they contribute to the evolution of an inclusive and biblio-diverse publishing ecosystem. In the second part I will detail ways to support community-driven open access journals, e.g., through substantial funding, coaching, and networking. Following-up on this, I will end with introducing a network developed by the Alexander von Humboldt Institute for Internet and Society that provides information materials and increases visibility for these journals.

1 INTRODUCTION

The open access movement dates to the mid-1990s and is widely understood as a direct reaction to the so-called "serial crisis" (see, e.g., Dobusch & Heimstädt, 2021, p. 430ff.) in academic publishing—this concerned a substantial and disproportionate increase in subscription costs, which led to affordability issues for public and academic libraries and a wave of journal subscription cancellations. This lack of access not only complicated the work of the academic community but led publishers to further increase subscription costs.

Advocates of the open access idea addressed this issue and used the opportunities offered by electronic publishing to loosen the grip of multi-corporate publishing enterprises on research dissemination. If research results, in the form of scholarly articles, can be distributed electronically, printing and publishing services can be rendered obsolete. Beyond these practical considerations, the rising open access movement, at least in its most radical manifestations, challenged the status quo of academic publishing and tried to advance towards more equitable financing and business models.

Two key documents of the open access movement outline their core demands and prospects. The seminal *Budapest Open Access Initiatives Declaration* from 2001 defines the availability and accessibility of research literature in the most extensive way and accepts limitations only "to give authors control over the integrity of their work" (BOAI, 2001). Adding to this, the 2003 *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities* understands open access as a "comprehensive source" within a web that ought to be "sustainable, interactive, and transparent" (Berlin Declaration, 2003).

In the following years, organizations that fund and conduct research successfully fostered a transition to open access and unlocked scientific articles, books, and data on an enormous scale. To give a few examples:

- There has been a steady growth in open access publications, accompanied by the growing popularity of open-source publishing tools and technologies (see the increasing use of editorial management systems like OJS, which has significantly contributed to the rise of community-driven publishing).
- National and international research funders have acknowledged the relevance of open access by requiring their grant recipients to publish their results—if possible—under open conditions (see the European Commission with Horizon 2020 or Horizon Europe respectively).
- National based research councils have also participated in the radical shift to openness and accessibility in the system of knowledge distribution: For example, recently, the German

Wissenschaftsrat released “Recommendations on the Transition of Academic Publishing to Open Access,” which is a strong statement on future policies in German academic publishing.

- International research consortia (e.g., cOAlition S) have issued policies and strategy papers in order to make open access quotas mandatory and provided transparent guidelines for better open practices in academic publishing.

Considering the above-mentioned developments, open access is soon to be the new standard in academic publishing—if this is not the case already—and promotes accessibility, findability, interoperability of research results and data. Ideally, it should help to create mutually dependent communities that care for and share knowledge.

Yet, in response to the growth and popularity of open access, publishing corporations have adapted their business strategies to generate new streams of revenue. Consequently, new business models have emerged that involve charging authors publication fees (called “article processing charges”) and created new inequalities and dependencies. In such cases, open access is not being used to equalize access but to enable new commercial means of knowledge dissemination. At the same time, and in the wake of the advancing digitalization of science, practices such as science tracking and predatory publishing have emerged and turned research data and articles into a mere currency within a larger system of knowledge dissemination. Broadly speaking, it seems that major publishing corporations remain largely in control of a significant portion of academic research despite not being sustainable, interactive, or transparent. This directly conflicts with the Berlin Declaration: Academic freedom, digital independence and digital sovereignty are being threatened by a commercial subversion of open access. So, the question is: How can researchers publish open access *and* remain in control of their article and data?

2 EXPLAINING COMMUNITY-DRIVEN PUBLISHING

One of the most intuitive strategies to counter the influence of the publishing industry (with its various attempts at commodification) is to exclude these commercial third parties and handle the publication and distribution of scholarly knowledge yourself. Often referred to as community-driven or scholar-led publishing, it is a practice that predates large-scale and commercial publishing—in fact, many of the most prestigious scholarly journals were founded by learned societies—but it has gained pace within the past ten years (cf. Morrison, 2016; Adema & Stone, 2017). While there is no proper definition of community-driven journals, science blogs, and book projects—and a large variety of them—there are a few common characteristics: Community-driven publishing endeavors are carried out on behalf or in the name of academia and academics (Moore, 2019); their day-to-day operation is based significantly on in-kind contributions or “gifted labor” by scholars (Adema/Moore, 2018, p. 8);

there are no charges for readers or authors in order to publish in or read them, which qualifies them for the diamond open access route (cf. Bosman et al., 2021); and most of them identify as nonprofit and noncompetitive in the broadest sense while emphasizing the common good and cooperation as their motivating factors. Based on these characteristics, community-driven publishing is an integral part of a diverse publishing ecosystem that fulfills two main functions.

1. Community-driven publishing projects foster a culture of experimental, collaborative and community-owned approaches to disseminating knowledge. This culture facilitates the creation of new output formats (beyond the somewhat dated peer-reviewed article), the development and testing of more inclusive processes of quality assurance, the revision of workflows, and the identification of administrative best practices etc.
2. Community-driven publishing projects enable self-determined and autonomous decision making in a time and age where consumers' and researchers' "digital sovereignty" is threatened (cf. Pohle/Thiel, 2020). More specifically, scholars remain largely in control over publishing (meta) data if they use open-source software and applications (in the form of Open Journal Systems). At the same time, they can question the widespread and nontransparent system of assessing impact through bibliometrics while increasing acceptance for other forms of evaluation, e.g., by alt metrics. Lastly, community-driven publishing projects may use license models that are approved for creating "free cultural works" (for instance, Creative Commons licensing).

Beyond these crucial functions for the open access ecosystem and its stakeholders, community-driven publishing seems to be guided by three principles that cater to the idea of a fair and truly open publishing landscape.

2.1 INCLUSIVE GOVERNANCE

To start with, to make community-driven publishing projects successful, initiators must mobilize and activate stakeholders by insisting on the common cause—that is, rebuilding the "broken" system of scholarly publishing. In doing so, they must come up with alternative ideas of "community governance"—e.g., around concepts like "mutual reliance," "care," and a variety of forms of "commoning" As Adema and Moore recently pointed out, "good governance requires rules and community trust within a social setting" (Adema/Moore 2020). Yet this somewhat entrepreneurial spirit should not be motivated by a fetish for technological innovation or the associated business around it but rather by the desire to create relations between projects and stimulate the bond within the community.

2.2 SCALING SMALL

Another part of this systemic change is to question the widespread economies of scale and the attempt to make any business venture "scalable." For academic publishing, stakeholders should consider "scaling small" (Adema & Moore, 2021), continuing catering to their niche, or, if any evolution is required, becoming more diverse—e.g., in terms of output format, audiences, quality assurance, impact, and distribution channels. While it is crucial for scholars to "stay in the market" of scholarly publishing, there are many steps that can be taken to ensure a resilient and robust structure of community-driven publishing, as Adema and Moore outline in their article (Adema & Moore, 2021). Stakeholders can, first, build horizontal support structures amongst like-minded publishing projects and therefore create a mutually reliable network of publishing partners. And they can, second, establish vertical collaborations—e.g., with funders, libraries, developers etc.—to create multi-stakeholder ecologies. This approach of horizontal and vertical networking is guided by collaboration instead of competition and reflects the aforementioned inclusive approach to governance.

2.3 CREATIVE FINANCING

While governance and business models define the internal structures and external relations, community-driven publishing must rethink and redefine funding and financing as well. David Ottina, one of the directors of the Open Humanities Press, once challenged the academic community to rethink scholarly communications and find an answer to "how we can make it resilient in the face of technological, institutional, and funding volatility" (Ottina 2013). In that regard, ideas from a recent, global-scale diamond open access study (Bosman et al., 2021) indicated that diversified income streams, constant public support, and common and open infrastructures are keystones for a robust architecture of fee-free scholarly publishing.

3 SUPPORTING COMMUNITY-DRIVEN OPEN ACCESS

As many studies have shown, community-driven publishing is essential for a diverse open access ecosystem. Yet many projects struggle or, even worse, cease operations (see scholar-led.network, 2021). But what can researchers, open access activists, developers and librarians do in order to support them?

In two research projects, the Alexander von Humboldt Institute for Internet and Society assessed the support that is needed to sustain community-driven publishing (Waidlein et al., 2021; Wrzesinski et al, 2021). In conclusion, they identified three areas in which this publishing segment requires substantial assistance.

3.1 FUNDING

In order to maintain high publication quality and stay a reliable partner for researchers, community-driven journals need long-term and robust financial support that is part of a coherent funding strategy by public and private stakeholders. Ideally, these funding structures and organizations should be led by the academic community so that the distribution of subsidies is guided by the interests of academia. As studies on the precarious situation of small and interdisciplinary journals have shown, funding also needs to be extended to the margins of open access publishing (Bosman et al., 2021).

3.2 COACHING

Academic editors and journal managers need dedicated information materials, guidelines, and peer-to-peer consulting that provides practical support for the day-to-day journal business. This will increase efficiency and streamline workflows, which in turn reduces administrative overheads and transaction costs. Additionally, external experts and publishing practitioners can assist journals in analyzing their business models and provide benchmarks for the editorial work.

3.3 NETWORKING

As a relatively new phenomenon, community-driven publishing needs an influential lobby that represents its interests in front of relevant organizations that fund and conduct research. Specifically, this includes creating awareness of the "gifted labor" and effort provided by editors and infrastructures and protecting smaller publishing projects as an integral part of a biblio-diverse publishing environment. Regular networking events are one way to go and offer opportunities to discuss developments and trends in community-driven publishing.

4 TURNING RESOURCES INTO PRACTICES

The project "Scholar-led Plus" at the Alexander von Humboldt Institute for Internet and Society (HIIG) is dedicated to a systemic change and addresses multiple of the above-mentioned challenges by developing a comprehensive set of knowledge resources. Considering the community of practice as a community of experts, the project has formed working groups on six key topics in community-driven publishing that were previously identified in several stakeholder meetings:

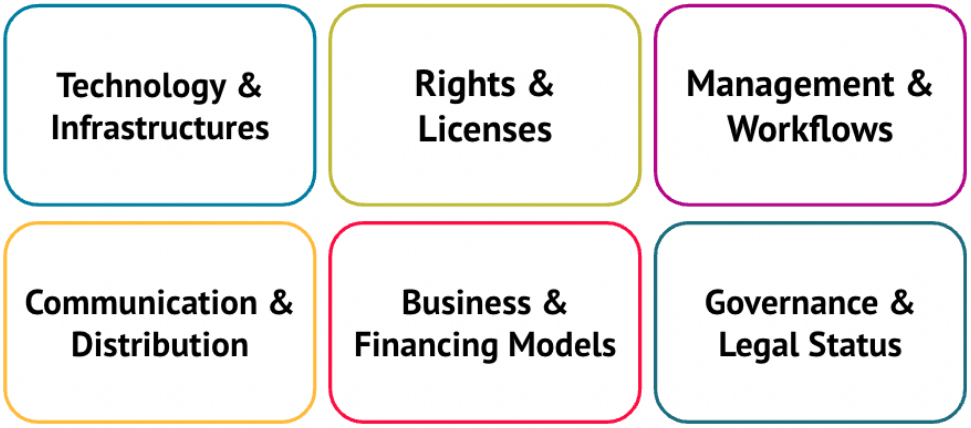


Figure 1. Key topics in community-driven publishing.

These working groups have worked to prepare six publication manuals and are serving as points of contact for practical inquiries on issues related to community-driven publishing in Germany. Towards the end of the project in Spring 2023, there will also be a set of (recorded) webinars that document the practical knowledge resources. Adding to this, HIIG will create strategic knowledge resources based on a Delphi study and a workshop series. This includes (1) a strategy paper on future trends and scenarios and (2) a policy paper on how to build resilient support for community-driven publishing. Both the practical and strategic resources will be made further accessible through roundtable discussions at HIIG in the first half of 2023, which will offer room for networking and exchange on the future of scholarly publishing.

5 REFERENCES

1. Adema J. & Moore S. A., (2021) “Scaling Small; Or How to Envision New Relationalities for Knowledge Production”, Westminster Papers in Communication and Culture 16(1), 27-45. https://doi.org/10.16997/wpcc.918
2. Adema, J., & Moore, S. A. (2018). Collectivity and collaboration: imagining new forms of communality to create resilience in scholar-led publishing. Insights, 31(0), 3. https://doi.org/10.1629/uksg.399
3. Adema, J., & Stone, G. (2017). The surge in New University Presses and Academic-Led Publishing: an overview of a changing publishing ecology in the UK. LIBER Quarterly: The Journal of the Association of European Research Libraries, 27(1), 97–126. https://doi.org/10.18352/lq.10210
4. Berlin Declaration (2003). Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. https://openaccess.mpg.de/Berliner-Erklaerung
5. Bosman, J., Frantsvåg, J. E., Kramer, B., Langlais, P.-C., & Proudman, V. (2021). OA Diamond Journals Study. Part 1: Findings. Zenodo. https://doi.org/10.5281/zenodo.4558704
6. Budapest Open Access Initiative (2001). Declaration of the Budapest Open Access Initiative. https://www.budapestopenaccessinitiative.org/read/
7. Dobusch, L. & Heimstädt, M. (2021). Strukturwandel der wissenschaftlichen Öffentlichkeit: Konstitution und Konsequenzen des Open-Access-Pfades. Leviathan Special Issue 37, 425-454.
8. Moore, S. (2019). Open *By* Whom? On the Meaning of ‘Scholar-Led.’ ScholarLed Blog, October 24. https://blog.scholarled.org/on-the-meaning-of-scholar-led/
9. Moore, S., & Adema, J. (2020). Community Governance Explored. COPIM. https://doi.org/10.21428/785a6451.20a5c646
10. Morrison, H. (2016). Small scholar-led scholarly journals: Can they survive and thrive in an open access future?: Small scholar-led scholarly journals: Can they survive and thrive in an open access future? Learned Publishing, 29(2), 83–88. https://doi.org/10.1002/leap.1015
11. Ottina, D. (2013). From Sustainable Publishing To Resilient Communications. TripleC, 11(2), 604-613. https://doi.org/10.31269/triplec.v11i2.528
12. Pohle, J. & Thiel, T. (2020). Digital sovereignty. Internet Policy Review, 9(4). https://doi.org/10.14763/2020.4.1532
13. scholar-led.network. (2021). Das scholar-led.network-Manifest. https://graphite.page/scholar-led-manifest/
14. Waidlein, N., Wrzesinski, M., Dubois, F., & Katzenbach, C. (2021). Working with budget and funding options to make open access journals sustainable. HIIG Discussion Paper Series 2021-1. https://doi.org/10.5281/zenodo.4558790
15. Wrzesinski, M., Riechert, P. U., Dubois, F., & Katzenbach, C. (2021). Working with publication technology to make open access journals sustainable. HIIG Discussion Paper Series 2021-2. https://doi.org/10.5281/zenodo.4558781

Proceedings of the Weizenbaum Conference 2022:
Practicing Sovereignty. Interventions for Open Digital Futures

OPEN HARDWARE AND SCIENTIFIC AUTONOMY IN GERMANY

HOW TRANSFER ACTIVITIES CAN BECOME MORE EFFECTIVE

Voigt, Maximilian
Open Knowledge Foundation Germany
Berlin, Germany
maximilian.voigt@okfn.de

KEYWORDS

open hardware; open science; knowledge transfer; technology transfer

DOI: 10.34669/wi.cp/4.9

1 INTRODUCTION

Technology is a fundamental instrument of human action. Those who develop and distributes it therefore have a major influence on the way we act. Hardware plays a central role in digital technologies as it is also the basis for all software. Since industrialization, knowledge about hardware has migrated more and more into the hands of the very few—through the division of labor and industrial processes. This has created dependencies (Simondon, 2012). With open hardware, people all over the world want to counteract this development and make technology more participatory. The aim is to make hardware understandable, repairable, and changeable. This includes open design, freely licensed documentation and project files, and the use of standard parts.

Science should play a central role in this context. Because universities generate knowledge financed by the publicly funded system. Scientists working on publicly funded projects are also engaged in hardware development, which is the basis for numerous innovations around the world that improve science itself. This makes them part of the innovation system, which should be accessible without barriers.

So, universities have the responsibility to transfer knowledge, an activity that has become increasingly important (Siegel & Wright, 2015). Many institutions make knowledge transfer requirements explicit in dedicated transfer strategies. The focus is on the dissemination of scientific knowledge to "society." Universities operate knowledge transfer offices for this purpose. However, they often miss their target. The notion of society is often narrowed down to economics. And transfer activities are often limited to start-up consulting and patenting activities (Nilsen & Anelli, 2016).

Open access and open science hardware are elementary factors in sharing knowledge more widely and effectively. They also form the basis for more autonomy in science. The present article will show this by means of examples. In addition, it will review aspects of a reformation of transfer offices. On this basis, university institutions can fulfill their responsibilities, increase the sustainability of research projects, and contribute to distributive justice.

1.1 WHAT IS OPEN SCIENCE HARDWARE - A SHORT DIGRESSION

Open hardware, often known as open-source hardware, is a technology-transfer approach in which hardware designs are made publicly available online for anyone to use, alter, and commercialize.
The often-quoted definition of the Open Source Hardware Association states:

> The hardware's source, the design from which it is made, is available in the preferred format for making modifications to it [...] Open-source hardware gives people the freedom to control their

technology while sharing knowledge and encouraging commerce through the open exchange of designs.[22]

Any piece of hardware used for scientific research that is available for purchase, assembly, usage, study, modification, sharing, and sale is referred to as "open hardware for science." It consists of both standard laboratory tools and auxiliary supplies such as sensors, biological reagents, and analog and digital electronic parts.

2 POTENTIAL FOR SCIENCE THROUGH OPEN SCIENCE HARDWARE

Around the world, there are committed scientists who advocate for open hardware in the science system. One network that has become known worldwide is the Gathering for Open Science Hardware network. People are committed because science itself often depends on technologies that have been specifically developed during the research process itself. Scientific measurement instruments are one example. They are often developed as part of research projects. Scientists involved then set up a company and sell the instruments back to scientific institutions. In this way, the knowledge about the instruments disappears from the institutions and dependencies arise. Service contracts may arise as a result, and while they can have advantages and save money (Wang & Richardson, 2020), they can also become problematic in the long run.

One example of this arises when companies go bankrupt or take products off the market, which often happens when company business models focus on servitization (Neely, 2008). Purchased devices then become unusable, because all the information and spare parts come from the manufacturer. A particularly tragic example of this arose in the field of biotechnology—the manufacturer Second Sight Medical Products withdrew its retinal implants from the market and people who had the implant in them had no way to maintain them.[23]

Another side of this issue is that such services are unaffordable for facilities in countries with lower socioeconomic development. As a result, they are unable to repair and maintain the equipment. "The World Health Organization estimates that 70% of donated medical equipment in sub-Saharan Africa is out of service" (Arancio & Shannon, 2022). Closed, outsourced technologies thus inhibit development, understandability, and reparability. In addition, they give rise to barriers that deprive socioeconomically disadvantaged countries of development opportunities. In contrast to proprietary approaches, approaches that develop such tools as open hardware have many advantages. There are numerous examples already.

[22] https://www.oshwa.org/definition/
[23] https://spectrum.ieee.org/bionic-eye-obsolete

2.1 AUTONOMY & COST SAVINGS THROUGH DEVELOPMENT OF OPEN INSTRUMENTS

Full-featured commercial systems are frequently unnecessary for those of modest means; instead, a straightforward open hardware solution could be adequate or even better. This is particularly true for lab instruction, where employing stripped-down open hardware may more effectively explain the fundamental measurement techniques than a closed-box commercial product. Open hardware's lower cost may also make it possible to give equipment on a "one-per-person" rather than "one-per-class" basis, improving the learning environment for students (Mello, 2020).

2.2 SUSTAINABILITY THROUGH COMMUNITY BASED DEVELOPMENT

Often, scientific projects end when public funding runs out. Open-source publication and community building can help to prevent this. "The long-term sustainability of a project such as [OpenFlexure] depends on the formation of a community, which is now active on the project's repositories on GitLab.com. As well as questions and bug reports, we have had contributions with fixes and improvements (...)" (Collins et al., 2020).

2.3 EFFECTIVE KNOWLEDGE TRANSFER AND INNOVATION PROMOTION

Experienced research institutions like CERN that have been publishing their knowledge and technologies open source for years have achieved much better reach for their transfer activities. In addition, they have also succeeded in stimulating innovations. As the following quote indicates:

> The Open Hardware Repository currently hosts more than 100 projects, ranging from small projects with a few partners to bigger projects with multiple contributors from both industry and academia. A dozen companies are actively involved in projects in the Open Hardware Repository, and some produce the physical hardware for CERN and other customers. CERN plays an important part also as a pilot customer for the hardware, legitimising the quality, making it easier for companies to sell it to other customers at a later stage. The Open Hardware Repository has led to an unprecedented re-use of existing design among scientific collaborators and internally at CERN. (Nilsen & Anelli, 2016)

2.4 EASIER ADAPTABILITY AND REPRODUCTION THROUGH MICROCONTROLLERS

The capacity to control and automate hardware is now more accessible than ever thanks to the development of robust-yet-user-friendly microcontroller and microprocessor platforms. The work of the instrument developer can be significantly simplified by the wealth of built-in capabilities included in modern microcontroller development boards. Timer functions for precise task scheduling, analog-

to-digital (ADC) converters for reading analog input signals, digital-to-analog (DAC) converters for creating arbitrary voltage waveforms, and hard-wired digital communication protocols for quick data exchange with other digital hardware are all useful features. Frequently, affordable add-on boards can be used to add missing capabilities regarding signal conditioning, motion control, wireless communication, and audio or image processing. This shifts functionality from the hardware to the software, thus simplifying the replication process (Mello, 2020; Fisher & Gould, 2012).

3 CHALLENGES IN THE ESTABLISHMENT OF OPEN SCIENCE HARDWARE

3.1 OPEN HARDWARE IS ABSENT FROM OPEN SCIENCE STRATEGIES OR DEFINITIONS

Despite numerous advantages of open hardware in the scientific context, there is still little development in this area. This becomes particularly evident when looking at various open science strategies published by scientific associations in Germany, like the *Wissenschaftsrat*.[24] Here, terms like "open-source hardware" do not even appear. Patents are also completely left out, even though they are an essential part of the scientific publication system. The German federal government even explicitly excludes patents. "The decision to exploit results commercially, e.g., by patenting, also remains unaffected."[25]

This shows how little attention is paid to open hardware and patents, even in the open science scene. This status is probably based on the proximity to economic interests pursued through technology transfer and an associated culture of intellectual property. A rethink is needed here, because, as will be shown in the following, technology transfer succeeds when it is pursued with open concepts.

3.2 THE PROBLEM WITH UNIVERSITY PATENTS

A challenge is the law on employee inventions in Germany. If employees have discovered something usable, they must report it to the responsible person at their research institution. This person then has the right to apply for a patent. Otherwise, the right is forfeited to the developer. A lot of resources go into this process. The aim behind this is to enable research institutions to attract third-party funding by entering into licensing agreements or selling the patents. For some institutes, the commercialization of their research is a way of obtaining extra income for the institute (Bray & Lee, 2000). However, more importantly, it is a way of strengthening the institute's attractiveness and role

[24] https://www.wissenschaftsrat.de/SharedDocs/Pressemitteilungen/DE/PM_2022/PM_0222.html
[25] https://www.bmbf.de/bmbf/de/forschung/digitale-wirtschaft-und-gesellschaft/open-access/open-access_node.html

in society. In addition, patent applications influence institutions' and individuals' reputations (Leitch & Harrison, 2005). The patent process is a major barrier to knowledge transfer and open science hardware, first, because the process ties up numerous resources, and second, because patent publications impose strong limits sharing and developing patented knowledge resources. There is the possibility of making patents compatible with open hardware, but this would require a fundamental change in the goals behind current practices.

A look at the figures shows that a rethink is needed. Patenting scientifically generated technologies is not sustainable, because most university patents do not even cover their costs of around €43,000 (Krause, 2017). Moreover, publications on university patents conclude that the supposed positive effects on society would likely have occurred even in the absence of patenting. "With the positive side effects described [in relation to society], however, it can be argued that the effects may also occur when inventions are published by universities but not patented. A causal link with the decision to apply for a patent for the invention is difficult to clearly establish" (Krause, 2017). Thus, the patent has no significant effect, but costs a lot of money.

These facts do not just apply to Germany. Even the USA, which is often cited as a positive example of scientific resource patenting, does not manage to monetize patents to any significant extent. Here, too, the positive effects are absent. "The results of applying this methodology to an average research university in the U.S. showed that it is not economic to invest in IP protection and patents." (Pearce, 2022). One study found that only 16% of knowledge and technology transfer offices in the United States were self-sustaining (Abrams et al., 2009).

4 OPEN HARDWARE AS AN OPPORTUNITY FOR TECHNOLOGY TRANSFER

The above-mentioned aspects show that there is a need for a rethink of the transfer activities of research institutions. Although it is common practice for universities in the United States and Germany to have transfer offices, it is not economically rational to continue to support them with their current alignment. "Instead, to increase the economic bottom line of the university as well as increase the good that university research does for society, universities should open source all their innovation." (Pearce, 2022)

4.1 THE ROLE OF TRANSFER OFFICES

Knowledge and technology transfer offices have been created in most universities and research centers to manage the dissemination process (Siegel & Wright, 2015). The activities focus particularly

on obtaining third-party funding and building reputation. Revenue generation is only a part of the picture, and knowledge and technology transfer offices have been found to increase access to external funding, to promote innovation and entrepreneurship, and to contribute to other public benefits (McDevitt et al., 2014). However, there are several ways in which this can be done successfully while simultaneously benefitting transdisciplinary communities and socioeconomically disadvantaged actors in addition to economic actors.

In this regard, Upstill & Symington (2002) argued that there are three basic modes for technology transfer from public research to the business sector:

- Noncommercial transfer: seminars, informal contacts, publications, secondments, and staff exchange and training
- Commercial transfer: collaborative research, contract research, consulting, licensing and sale of intellectual property and technical services
- New company generation: direct spin-offs, indirect spin-offs, and technology transfer companies

Noncommercial transfer, such as publications, presentations, and informal exchanges, have been found to be among the most important ways to diffuse public research to industry (Cohen et al., 2002). Even in institutes known for their large patent output, such as MIT, publications outnumber patents as a mean of transferring knowledge (Agrawal & Henderson, 2002). To increase the impact of their research and developments, organizations should make their knowledge available free through open-source licenses and other open mechanisms (Sorensen & Chambers, 2008). CERN, as one of the leading research institutions in the field of science and technology, has extensively analyzed its transfer activities. They concluded that pursuing open approaches, such as free licensing of technical developments, produced a far-reaching impact for their transfer goals (Nilsen & Anelli 2016).

So, instead of focusing on technology transfer by patenting and licensing technological developments, institutions of this kind should publish knowledge in freely licensed publications, for example, via portals such as the *Journal of Open Hardware*.

4.2 Measures for the implementation of Open Science Hardware

Transfer offices are important factors for the successful establishment of open science hardware. In a sense, they already provide the infrastructure needed to make science more open. All that is needed is a restructuring of their work. Suggestions for this have been provided, for example, by the international GOSH network.[26] These recommendations include:

[26] https://openhardware.science/policy-briefs/

- Transfer offices should provide advice after an invention disclosure regarding open hardware and not support patenting in the first place.
- Instead of patent applications, technical developments should be extensively documented and published.[27] Support is needed for this, because documenting hardware requires special procedures, accuracy, and compliance with standards. One guideline for structuring good open documentation is DIN SPEC 3105-1. Transfer offices could support researchers and guide them through the documentation process.
- Transfer offices should also help build and maintain developer communities to make research projects with a focus on technology more sustainable. Building a developer community increases the likelihood that developments will continue to be worked on and used after the funding expires.
- Working with such open communities and producing open-source knowledge also requires dedicated skills. Therefore, competence building in the application and development of open-source tools is needed. This also includes teaching the basics of intellectual property law and collaborative work, for example, by conceptually designing courses with the re-usability of their results in mind.

5 CONCLUSION

Opening up universities (open science) and sharing scientific resources under a free license (open access) are important trends of our time (Morais et al., 2021). This increases trust in science, makes science more accessible, and improves scientific processes (Hyunjin et al., 2022). However, the focus on technical knowledge is lacking, especially in Germany. The patent as a publication form has not been the subject of open access strategies, although patents are an important part of the publication system. This leaves out knowledge about hardware and thus the development of scientific tools. But these are an important part of the science and innovation system. Publishing as open-source hardware could benefit these systems and reach a broader audience. This requires a cultural change on intellectual property and the reformation of transfer offices at universities. These should put fewer resources into patenting and licensing scientific inventions. Instead, they should support scientists in building developer communities and in creating and publishing open documentation.

[27] The publication “Open-Source Photometric System for Enzymatic Nitrate Quantification” shows how this could be done: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0134989

6 REFERENCES

1. Abrams, I., Leung, G., Stevens, A.J. (2009). How are US technology transfer offices tasked and motivated-is it all about the money. Research Management Review, 17 (1), 1-34.
2. Abrams, I., Leung, G., Stevens, A.J. (2009). How are US technology transfer offices tasked and motivated-is it all about the money. Research Management Review, 17 (1), 1–34.
3. Agrawal, A., Henderson, R. (2002). Putting Patents in Context: Exploring Knowledge Transfer from MIT. Management Science, 48 (1), 44–60.
4. Arancio, J., Shannon, D. (2022). Bringing Open Source to the Global Lab Bench. Issues in Science and Technology 38 (2), 18–20.
5. Bray, M.J., Lee, J.N. (2000). University revenues from technology transfer: licensing fees vs. equity positions. Journal of Business Venturing, 15 (5-6), 385–392.
6. Cohen, W.M., Nelson, R.R., Walsh, J.P. (2002). Links and impacts: the influence of public research on Industrial R&D. Management Science, 48 (1), 1–23.
7. Collins, J.T., Knapper, J., Stirling, J., Mduda, J., Mkindi, C., Mayagaya, V., Mwakajinga, G.A., Nyakyi, P.T., Sanga, V.L., Carbery, D., White, L., Dale, S., Jieh Lim, Z., Baumberg, J.J., Cicuta, P., McDermott, S., Vodenicharski, B., Bowman, R. (2020). Robotic microscopy for everyone: the OpenFlexure microscope. Biomed Opt Express, 11(5), 2447–2460.
8. Fisher, D., Gould, P. (2012). Open-Source Hardware Is a Low-Cost Alternative for Scientific Instrumentation and Research. Modern Instrumentation, 1 (2), 8–20.
9. Hyunjin S., David M. M., Samuel H. T. (2022). Trusting on the shoulders of open giants? Open science increases trust in science for the public and academics, Journal of Communication, Volume 72 (4), 497–510.
10. Krause, M. (2017). Erfolg Patentierter Hochschulerfindungen. Hürth: Carl Heymanns.
11. Leitch, C.M., Harrison, R.T. (2005). Maximising the potential of university spin-outs: the development of second-order commercialisation activities. R&D Management, 35 (3), 257–272.
12. McDevitt, V.L., Mendez-Hinds, J., Winwood, D., Nijhawan, V., Sherer, T., Ritter, J.F., Sanberg, P.R. (2014). More than money: the exponential impact of academic technology transfer. Technology & Innovation, 16 (1), 75–84.
13. Mello, J.d. (2020). Opening Up Instrumentation. The Chemical Engineer, 952.
14. Morais R., Saenen B., Garbuglia F., Berghmans F., Gaillard V. (2021). From principles to practices: Open Science at Europe's universities. 2020–2021 EUA Open Science Survey results. Brussels & Geneva, European University Association.
15. Neely, A. (2008). Exploring the financial consequences of the servitization of manufacturing. Oper Manag Res, 1, 103–118.
16. Nilsen, V., Anelli, G. (2016). Knowledge transfer at CERN. Technological Forecasting and Social Change, 112, 113–120.
17. Pearce, J. M. (2022). Full Cost Accounting Shows the Emperor Has No Clothes: Universities Investing in Technology Transfer via Patenting Lose Money. https://doi.org/10.5281/zenodo.6345120

18. Siegel, D. S., Wright, M. (2015). University Technology Transfer Offices, Licensing, and Start-Ups. The Chicago Handbook of University Technology Transfer and Academic Entrepreneurship. Chicago, USA.
19. Simondon, G. (2012). Die Existenzweise technischer Objekte. Zürich: diaphanes.
20. Sorensen, J.A.T., Chambers, D.A. (2008). Evaluating academic technology transfer performance by how well access to knowledge is facilitated – defining an access metric. The Journal of Technology Transfer, 33, 534–54.
21. Upstill, G., Symington, D. (2002). Technology transfer and the creation of companies: the CSIRO experience. R&D Management, 32 (3), 233–239.
22. Wang, Y., Richardson, D. S. (2020). To buy or to lease. EMBO Reports 21 (5).

Proceedings of the Weizenbaum Conference 2022:
Practicing Sovereignty. Interventions for Open Digital Futures

COVID-19 FROM THE MARGINS

NARRATING THE COVID-19 PANDEMIC THROUGH DECOLONIALITY AND MULTILINGUALISM

Masiero, Silvia
University of Oslo
Oslo, Norway
silvima@ifi.uio.no

Milan, Stefania
University of Amsterdam
Amsterdam, the Netherlands
s.milan@uva.nl

Treré, Emiliano
Cardiff University
Cardiff, UK
treree@cardiff.ac.uk

KEYWORDS

COVID-19; margins; decoloniality; multilingualism; big data from the South

DOI: 10.34669/wi.cp/4.10

ABSTRACT

Born as a multilingual blog in May 2020, *COVID-19 from the Margins* has offered a space for authors to voice the silenced narratives of the COVID-19 pandemic in any language chosen and representing multiple South(s) of the world (Milan & Treré, 2019). The blog became an open-access book in February 2021, and since then it has travelled across the globe to bring to light narratives of devoiced groups during COVID-19, generating debate on stories narrated by, amongst others, forced migrants, gig workers, ethnic minorities, people in economic poverty, and survivors of domestic violence. The project is divided into five sections—"Human Invisibilities and the Politics of Counting," "Perpetuated Vulnerabilities and Inequalities," "Datafied Social Policies," "Technological Reconfigurations in the Datafied Pandemic," and "Pandemic Solidarities and Resistance from Below"—which together contribute to the decolonial, multilingual project of narrating the COVID-19 pandemic through the voices of the systematically silenced. In this short paper, we reflect on the *COVID-19 from the Margins* experience and on its meaning towards a decolonial, multilingual narration of the COVID-19 pandemic.

1 INTRODUCTION

COVID-19 is the first pandemic of the datafied society. Over the last two years and a half, public and governmental narratives of the pandemic have been centered on quantifying the impact of the virus, counting and measuring its effects through data made available in different forms. Statistics and tabulations have told the official stories of the pandemic, visualizing it and mapping its diffusion across countries and regions. Official statistics have become the hegemonic narrative of the pandemic, and the route through which pandemic narratives were reported on and displayed.

And yet, not all voices have been equally heard or represented. As much as it was a story of visualized statistics, COVID-19 holds invisible narratives: these come from the groups, communities and individuals subject, in various ways, to systematic silencing in public discourse. These narratives come from migrants, informal workers, economically poor people, ethnic minorities, workers of the gig economy, and survivors of domestic violence, who, in different parts of the world, suffered the effects of the COVID-19 pandemic without being given a voice in it. This reproduces dynamics of systemic silencing at the margins, conceptualized, following Rodríguez (2017), as "a shortcut to speak of complex dynamics of power inequality."

2 THE PROJECT

Against this backdrop, our project aims to voice the systematically silenced narratives of the COVID-19 pandemic. Programmatically titled *COVID-19 from the Margins*, the project has created a virtual space where such narratives have a voice, adopting a multilingual format that allows the writers to use the language(s) they prefer to express their voices. Multilingualism, seen as a route to escape the constraints of the use of English as a lingua franca in academic settings (Suzina, 2021), is integral to the voice-giving purpose of *COVID-19 from the Margins*, and has characterized the project since its inception. In *COVID-19 from the Margins*, multilingualism is, in addition, a route to a decolonial approach, understood following Escobar (2018) as an approach that unhooks the production of knowledge from predominantly Western science and paradigms. Multilingualism and decoloniality, intertwined and necessary for each other, are the two pillars on which our project has been founded (Milan et al., 2021).

The project started off in May 2020 as a blog, which invited writers to provide narrations of the datafied pandemic from the margins. It was conceived, again following Rodríguez (2017), beyond geographical and geopolitical borders. Rather than offering a dichotomic vision of the global North and South (Pansera, 2018), we embraced a vision of the world's South(s) as a plurality of loci of oppression and resistance, a pluralism recognized since the early work of Milan and Treré (2019).

Based on the blog, which attracted contributions from a plurality of Souths and a multitude of settings of pandemic invisibility, resistance, and solidarity, we created an open-access book, relying on 47 of the multilingual contributions submitted and published in the blog. Embracing multiple settings, forms of oppression and routes to resistance in the datafied pandemic, the book is articulated into five sections—"Human Invisibilities and the Politics of Counting," "Perpetuated Vulnerabilities and Inequalities," "Datafied Social Policies," "Technological Reconfigurations in the Datafied Pandemic," and "Pandemic Solidarities and Resistance from Below"—which, taken together, offer a comprehensive picture of silenced histories from the datafied pandemic.

3 THE BOOK'S SECTIONS

Each section in the book illuminates a part that is integral to the narration of COVID-19 from the margins. In Section 1, "Human Invisibilities and the Politics of Counting," people from communities that suffer from systematic exclusion from mainstream narratives voice their version of the pandemic's events. These narratives, which the team received from the early days of the blog, noted how disease surveillance in the pandemic determined invisibility, raising concerns about justice in population data management. Coming from multiple locations and experiences, these narratives found a common matrix in problematizations of data justice, meant, following Taylor (2017, p. 1), as "fairness in the way people are made visible, represented and treated as a result of their production of digital data". In fact, multiple narratives in this section offer proper instantiations of data injustice, which acquires at least three forms according to Masiero and Das (2019):

- Legal injustice, when universal rights (e.g., to food and shelter) are made conditional on datafication, for example by enrollment in biometric databases;
- Informational injustice, where inaccurate or incomplete information is given to people about the way their data is used or will be used;
- Design-related injustice, where artefacts, for example contact tracing apps, are designed in such a way to produce harm in the individual (Costanza-Chock, 2020).

In Section 2, "Perpetuated Vulnerabilities and Inequalities," authors examine the crystallization of previous forms of inequality and vulnerability and the production of new ones in datafied forms during the pandemic. In one of the chapters, it is noted that, as "staying at home" became the new normal during the pandemic, LGBTQ+ communities, suffering ingrained social prejudice, may not have had a home to stay in. Several contributions, focusing on work in the gig economy, noted how gig workers were constantly at the forefront of risk in the pandemic while at the same time suffering the same perpetuated subalternity and subjection to the "rating economy" experienced before (Anwar

& Graham, 2021). As a whole, the section notes how pandemic technologies reified and reinforced existing subalternities, generating new inequalities on top of existing ones.

In Section 3, "Datafied Social Policies", authors explore multiple instances of social policies adopted during the COVID-19 pandemic. The sections contain, on the one hand, multiple instantiations of informational injustice as defined in Masiero and Das (2019): these pertain to the opaque cross-checking of citizen data across databases such as land registries, property, and population datasets, making the assignation of subsidies extremely uncertain and experimental (Cerna Aragon, 2021; López, 2021). On the other hand, the section illuminates the issue residing in the conditionality of essential social provision to digital authentication, which, where it fails, leaves needful users exposed to stark magnifications of the pandemic's risks. The section illuminates, overall, the risks associated with the datafication of social protection in the context of the COVID-19 pandemic, illustrating regularities across countries and reflecting on how such issues can be overcome.

In Section 4, "Technological Reconfigurations in the Datafied Pandemic," the focus shifts on how existing technologies have been repurposed during COVID-19. Contributions show how existing social media platforms have afforded new forms of collective action, while new safeguards have been added to data policies due to the vulnerability of communities to unfair data treatment. This problem, as the contributions of Das (2021) and Raghunath (2021) note, was exacerbated in those contexts where contact tracing apps were implemented without an adequate data protection framework, requiring a reconfiguration of existing data protection structures. Open data has been proposed as a route to overcome the issue and acts as a route to liberation from situations of outright pandemic negationism and oppression (Fussy, 2021).

In this spirit, the last section, "Pandemic Solidarities and Resistance from Below", illuminates new, datafied forms of solidarity and resistance that emerged, across countries and contexts, during COVID-19. Digital platforms, as a relatively new actor in the landscape of solidarity-making, acquired an important role as sites of contestation and also organization of solidarity across groups. Novel practices, such as citizen sensing, have raised important questions about how existing logics of resistance can be rapidly reorganized in emergency situations. Ending the book on a note of hope, the final section illuminates how technology can be reappraised against oppression, to engage the very practices of silencing that the whole project was designed to respond to.

4 A LOOK TO THE FUTURE

In its entirety, the *COVID-19 from the Margins* project is a whole that transcends its outputs (the blog; the book; its presentations) and seeks to initiate a movement that views the pandemic from the

world's multiple South(s), in line with the *Big Data from the South* project launched by Milan and Treré (2019). As a result, the project seeks to travel between the physical and virtual world to interact with new communities and audiences, breaking the barriers of traditional academic discourse and engaging domains of practice, activism, and civil society. This is what led us to propose the project for presentation in an interactive, academic-artistic forum like the Weizenbaum Conference, where we seek to build further interaction in our multilingual and decolonial venture. We believe that the transdisciplinary platform offered by the conference, by virtue its interactivity and cross-disciplinarity, can be the basis to further grow in terms of the activist potential that the project can generate.

Among the conference tracks, we saw the project contributing to the track on Datafication and Democracy as its themes—in terms of the politics of data, with its implications for counting and surveillance capitalism—overlapped directly with stories featured in the book. We believe that bringing *COVID-19 from the Margins* to track 2 of the conference has been the start of an important conversation, leading to further abating of the barriers that silence particular communities and further building of the power of cross-disciplinarity in telling the silenced stories of the datafied pandemic.

5 REFERENCES

1. Anwar, M. A., & Graham, M. (2021). Between a rock and a hard place: Freedom, flexibility, precarity and vulnerability in the gig economy in Africa. *Competition & Change*, 25(2), 237–258.
2. Cerna Aragon, D. (2021). On not being visibile to the state: The case of Peru. In Milan, S., Treré, E., & Masiero, S. (Eds.), *COVID-19 from the Margins: Pandemic Invisibilities, Policies and Resistance in the Datafied Society* (pp. 120–125). Amsterdam: Institute of Network Cultures.
3. Costanza-Chock, S. (2020). Design justice: Community-led practices to build the worlds we need. Boston: The MIT Press.
4. Das, S. (2021). Surveillance in the time of coronavirus: The case of the Indian contact tracing app Aarogya Setu. In Milan, S., Treré, E., & Masiero, S. (Eds.), *COVID-19 from the Margins: Pandemic Invisibilities, Policies and Resistance in the Datafied Society* (pp. 57–61). Amsterdam: Institute of Network Cultures.
5. Escobar, A. (2018). Designs for the Pluriverse. New York: Duke University Press.
6. Fussy, P. (2021). Liberating COVID-19 Data with Volunteers in Brazil. In Milan, S., Treré, E., & Masiero, S. (Eds.), *COVID-19 from the Margins: Pandemic Invisibilities, Policies and Resistance in the Datafied Society* (pp. 241–245). Amsterdam: Institute of Network Cultures.
7. López, J. (2021). The case of the Solidarity Income in Colombia: The experimentation with data on social policy during the pandemic. In Milan, S., Treré, E., & Masiero, S. (Eds.), *COVID-19 from the Margins: Pandemic Invisibilities, Policies and Resistance in the Datafied Society*. (pp. 126–128). Amsterdam: Institute of Network Cultures.
8. Masiero, S. (2022). Decolonising critical information systems research: A subaltern approach. *Information Systems Journal*, online 8 July 2022.
9. Masiero, S., & Das, S. (2019). Datafying anti-poverty programmes: Implications for data justice. *Information, Communication & Society*, 22(7), 916–933.
10. Milan, S., Treré, E., & Masiero, S. (2021). COVID-19 from the Margins: Pandemic Invisibilities, Policies and Resistance in the Datafied Society. Amsterdam: Institute of Network Cultures.
11. Milan, S., & Treré, E. (2019). Big data from the South (s): Beyond data universalism. Television & New Media, 20(4), 319–335.
12. Pansera, M. (2018). Frugal or fair? The unfulfilled promises of frugal innovation. Technology Innovation Management Review, 8(4), 6–13.
13. Raghunath, P. (2021). COVID-19 and non-personal data in the Indian context: on the normative ideal of public interest. In Milan, S., Treré, E., & Masiero, S. (Eds.), *COVID-19 from the Margins: Pandemic Invisibilities, Policies and Resistance in the Datafied Society* (pp. 200–202). Amsterdam: Institute of Network Cultures.
14. Rodríguez, C. (2017). Studying media at the margins: Learning from the field. In Pickard, Victor & Yang, Guobin (Eds.), Media activism in the digital age (pp. 49–60). London: Routledge.
15. Suzina, A. C. (2021). English as lingua franca. Or the sterilisation of scientific work. Media, Culture & Society, 43(1), 171–179.

16. Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. Big Data & Society, 4(2), 1–14.

Proceedings of the Weizenbaum Conference 2022:
Practicing Sovereignty. Interventions for Open Digital Futures

AUTOFICTIONAL DOCUMENTARY, SITUATED KNOWLEDGES, AND COLLECTIVE MEMORY

ON DEAR CHAEMIN (2020)

Bae, Cyan
Leiden University
Leiden, the Netherlands
c.bae@fsw.leidenuniv.nl

KEYWORDS

autofictional documentary; citational narrativizing; situated knowledge; collective memory; COVID-19

DOI: 10.34669/wi.cp/4.11

ABSTRACT

The COVID-19 pandemic has disproportionately affected communities already marginalized in pre-coronavirus societies, aggravated by socio-political technologies of racialization, sexism, homo- and transphobia. *Dear Chaemin* (directed by Bae, 2020) is an autofictional documentary series of three video letters sent from The Hague to the director's sister in Seoul amid isolation. The film juxtaposes the Korean and Dutch contexts of state surveillance, entangled with the b/ordering technologies against queer communities in Seoul and Asian communities in Europe. This paper explores autofictional documentary as an audiovisual method to engage with contemporary dynamics of international politics. First, I summarize the arguments made in the three chapters of the film *Dear Chaemin*. Second, I propose autofictional documentary as an effective cinematic mode that accounts for situated knowledges and critiques collective memories. Finally, I explore how the autofictional mode is further contextualized through the use of unconventional, non-lens-based audiovisual material.

1 INTRODUCTION

The COVID-19 pandemic has disproportionately affected communities that were already marginalized in pre-coronavirus societies, aggravated by socio-political technologies of racialization, sexism, homophobia, and transphobia. Philip di Salvo (2021) provides an overarching account of how the pandemic recodified the everyday politics in the "datafied pandemic." The chapter demonstrates how solutionism (i.e., the idea that technology can solve all social and political problems) and surveillance became ever more intertwined with the political efforts to mitigate the spread of the coronavirus worldwide. With the technologized interventions, preexisting forms of oppression and violence intensified as these measures brought about uneven consequences. These issues urgently call for critical readings of the worldwide crisis. Such tasks should depart from the quantification and datafication of the harm that took place.

Dear Chaemin (directed by Bae, 2020) is an autofictional documentary series of three video letters sent from The Hague to the director's sister in Seoul amid isolation. The film juxtaposes the Korean and Dutch contexts of state surveillance, entangled with the b/ordering technologies against queer communities in Seoul and Asian communities in Europe. This project questions the normativity embedded in the (techno-)optimistic outlook toward the post-Corona futures.

This paper explores autofictional documentary as an audio-visual method to engage with contemporary dynamics of international politics. First, I reiterate the arguments made in the three chapters of the film *Dear Chaemin*. Second, I propose autofictional documentary as an effective cinematic mode that accounts for situated knowledges and critiques collective memories. Finally, I explore how the autofictional mode is further contextualized through the use of unconventional, non-lens-based audio-visual material.

2 FOR "BEARING THE WORSENING": *DEAR CHAEMIN*

The first letter serves an introductory role in the series. It traces the internationally praised South Korean measures to contain the coronavirus back to the history of the Resident Registration Numbering (RRN) system. South Korea's technology-intensive efforts foregrounded fine-grained locational data and social network analysis to track and target individuals for containment and treatment (Ramraj, 2021; French and Monahan, 2020). While deemed exceptional and temporary, the aggressive measures were only made possible through the infrastructural RRN system and its inherently gendered governance of citizenship. South Korea's RRN system assigns a 13-digit number to all residents in the country to manage and control the population. Established during the Park Chung-hee dictatorship in 1968, RRN constituted the institutional foundation for state security from

the threat of North Korea (Hong, 2007, p. 324). The seventh digit, which follows the first six digits signifying the birthdate of the holder, indicates the individual's binary gender assigned at birth. As the numbering system is implemented in every aspect of living in South Korea—education, banking, employment, law enforcement, and healthcare—the institutionalized body is continually subject to the cisnormative order. The coronavirus measures and policies, ironically yet unsurprisingly, jeopardize queer, trans, and gender non-conforming people's health through the infrastructural technology of the RRN system (Lee et al., 2021; Park et al., 2021).

The second letter investigates the racialized "lateral surveillance" (Andrejevic, 2004) in the Dutch context. Lateral surveillance, or peer-to-peer monitoring that happens among individuals rather than surveillance by institutions, is known to target the racialized other disproportionately. The events following the outbreak of the COVID-19 in Wuhan, China, again exemplified how the racialized other is perceived as destabilizing and threatening to the white body politic (French and Monahan, 2020, p. 5). This chapter represents an autoethnographic account of the racialized experience in Europe amid lockdowns. Sinophobia during the COVID-19 pandemic not only operated according to citizenship but also extended to the East Asian-looking population. Furthering the arguments made in the first letter, this chapter also discusses how the racialized migrant experience is complicated by queerness. The Western homonationalist agenda extends its reach to non-Western queer beings. Homonationalism, or the co-optation of queer identity claims by nationalist discourses, regulates sexual-racial populations in terms of national identity (Puar, 2007). Promoting queer rights in European states, including marriage equality, registered partnership, and gender affirmation, does not only contribute to the branding of national homosexuality within borders but also generates a yearning for departure for non-European queers. When migration seemingly enables the queer migrant to achieve this desire, they face continued exclusion as they are denied the (white) queer citizenship by racializing logic.

The third and last letter examines the techno-solutionist outlook towards the post-COVID era through the figuration of the immune citizen. While earlier chapters present archival footage to evidence the proposition that state and interpersonal surveillance was not newly adopted for the public health emergency but was just the latest iteration, this chapter draws on archival footage to discuss the idealization of the safe, immune, and unthreatening body with reference to the exclusion of the contagious other. The collective imagination—which was strongest during the first year of the outbreak—and the anticipation of clear-cut emancipation from the disease, is critically assessed. Finally, this chapter draws on trans activist-scholar Ruin's (2019) work to question the temporal imagination of linear advancement. In her essay "Deo Naeun Miraeraneun Chakgak [The Illusion of a Better Future]," Ruin writes that "regardless of the hopes and aspirations for a linear improvement,

life and movements flow," as she examines archival material about South Korean queer history. The subtitle of the essay section, "Bearing the Worsening," encapsulates one of the most powerful lessons for trying times from the tradition of feminist and queer thinkers. The collective imagination of the post-COVID era must reject the romanticization of the pre-COVID era and instead embrace "a politics of epistemological humility" (Eng, Halberstam, and Muñoz, 2005). This premise honors the very fact that we are entering into an unknown time and space entangled with and embracing the uneven distribution of impossibilities, instabilities, and insecurities (Jun, 2021). A collective imagination of such politics foregrounding the queer epistemology will strive for a world-building for the otherwise that addresses these challenges while "staying with the trouble" (Haraway, 2016).

3 "THERE WERE NO LIES": AUTOFICTIONAL DOCUMENTARY AND COLLECTIVELY SITUATED KNOWLEDGES

"I know this letter will live short, and that these words will fade. Still, I wrote for memory. Half of this was a confession, and the other half a novel. There were no lies." (*Dear Chaemin*, 2020)

Autofictional approaches in documentary filmmaking have been identified as a sub-genre of documentary that emerged from the postmodern perspective with the demise of the "metanarrative" or "grand narrative" (Lyotard 1984). In autofictional documentaries, filmmakers "felt free to not only put themselves in their films, but to employ a wide range of stylistic approaches in telling their stories, and to maybe even lie a little" (Corbett, 2016, p. 52). Following the tradition of autofiction as a literary genre, autofictional filmmaking is understood as "a contemporary cinematic mode that challenges, and at times subverts, the generic limits of documentary and fiction film from a self-reflexive position" (Forné and López-Gay, 2022, p. 228).

I propose that autofictional documentary is an effective cinematic mode that accounts for situated knowledges and critiques collective memories. For *Dear Chaemin*, the autofictional mode was a strategic decision to re-stage the rapidly unfolding events of intertwined racism and queerphobia that followed the intensifying pandemic on a global scale. The challenges of its research and production lay, on one hand, in the racialization of the coronavirus targeted at Asian diasporas across countries and continents, while I was confined to the regional quarantine at the time of production. On the other hand, the tasks of this film included representing the experiences of queer and trans communities of color without reducing them to testimonial appropriations. Autofictional filmmaking, characterized by the aesthetics of ambiguity, allows a translation that "can seize on the same facts and events, but assembles them in a radically altered presentation, disorderly or in an order, which deconstructs and reconstructs the narrative according to its own logic with a novelistic design of its own" (Doubrovsky, 2013).

Autofictional documentary is particularly useful for narrativizing situated knowledge from a located position while speaking nearby different positionalities. Donna Haraway's (1988) theory on situated knowledges based on the embodied nature of all vision cannot be ignored when considering documentary filmmaking as a mode of knowledge production. While vision is taken as a metaphor to discuss scientific objectivity Haraway's the seminal work, filmmaking as a visualizing technology evidences that every perspective is partial and contextualized by its embodied standpoint. Through the characterized presence of the filmmaker, autofictional documentary enables a storytelling of "the view from a body, always a complex, contradictory, structuring, and structured body, versus the view from above, from nowhere, from simplicity" (Haraway, 1988). For Haraway, a mere partiality does not suffice. It is "the joining of partial views and halting voices into a collective subject position" (Haraway, 1988) that allows translations and solidarities. While Haraway emphasizes conversations between partial perspectives, however, the question of *how* these connections could be accessed is less explored (Jeong, 2013).

When it comes to filmmaking, the question of how to mobilize multiple partialities may align with Trinh T. Minh-ha's approach of "speaking nearby" instead of "speaking about." To speak nearby is:

> to acknowledge the possible gap between you and those who populate your film: in other words, to leave the space of representation open so that, although you're very close to your subject, you're also committed to not speaking on their behalf, in their place or on top of them. You can only speak nearby, in proximity (whether the other is physically present or absent), which requires that you deliberately suspend meaning, preventing it from merely closing and hence leaving a gap in the formation process. (Balsom, 2018)

For Trinh, it is an "attitude in life, a way of positioning oneself in relation to the world," rather than a technique or a statement that is materialized in all aspects of the film with a challenge to approach truth indirectly (Chen, 1992, p. 87). Filmmaker An van Dienderen (2017) interprets the approach of speaking nearby as ambiguity and hybridity aimed through fictional and archival filmic strategies. Embracing that documentary is a "flow between fact and fiction" (Trinh, 1990), Trinh's practice of speaking nearby entails a critique of objectivity that destabilizes power relations in knowledge production, namely the hierarchy between the filmmaker and the subject.

The autofictional extends the provocation—or the play between fact and fiction—to encompass a further malleability. For *Dear Chaemin*, this malleability in rearranging the reality is what enabled staging and performing dialogues between my situated knowledge and other positionalities without a goal for universal or journalistic objectivity. I do not wish to disclose what was fact or what was fiction in this film—not in this paper nor elsewhere. It is precisely through the

ambiguity of the autofictional that I could map the socio-political entanglements in proximity. In this project, I deploy citational narrativizing as the primary strategy that weaved a web of relations to speak nearby. I mobilize dispersed lived experiences based on news articles, testimonials, blog posts, and literary works into a community of lovers, friends, and families. Through scriptwriting in-between fact and fiction, news and novels, or documentary and narrative, I sought to collectivize partial and localized experiences to be contextualized in a form that promotes intimacy.

Two years after its production, my reflections on *Dear Chaemin* are also a recognition of it as an attempt to archive the early stage of the pandemic from a queer-feminist perspective. Forné and López-Gay (2022) conceptualize the autofictional as an archival practice of memorialization through a self-reflexive process. The ambiguity established between the filmmaker and audience also "suffuses the poetics of memory that [the autofictional films] deploy" (Forné and López-Gay, 2022,p. 229) and invites the viewer to actively interpret and participate in the process of memory construction. For *Dear Chaemin*, the collective memory it constructs primarily interrogates the normative conception of the pandemic in its earlier stage as a universal public health emergency. Further, this project leads to a questioning of the romanticization of the "old normal" as it rearranges historical traces.

In this vein, the autofictional mode can be further contextualized through the use of unconventional, non-lens-based audio-visual material. The research and development of *Dear Chaemin* in its pre-production were populated by news coverage and archival material, most of which were accessed online. Employing desktop recordings and 3D mapping, the film not only adheres to the project of documenting but also interrogates digital representations of marginalized experiences depicted online and in the mass media. It historicizes the securitization of the pandemic mediated by socio-technical assemblages of governance. As a result, *Dear Chaemin* becomes an archive that creates and examines both digital and analogue forms of individual and collective memories, which together shape contemporary memorialization processes.

4 CONCLUSION

In this paper, I have discussed how autofictional documentary as a sub-genre of documentary enables narrativizing that collectivizes differently located situated knowledges. I locate the malleability of the autofictional in reassembling realities as a cinematic strategy for speaking nearby other positionalities while embracing and advocating for the partiality of vision. Further, autofictional filmmaking—with the viewer—participates in the process of memory construction. Accompanied by non-lens-based material, the autofictional mode can engage with the socio-technical process of memorialization.

Through this paper and the film *Dear Chaemin*, I propose that autofictional documentary is a cinematic mode of research of studying contemporary dynamics of international politics that refuses to engage in all-knowing claims. It enables the filmmaker/researcher to present an embodied process of research with a "symmetrical embrace of (fluid and constantly shifting) alphabetical and material-aesthetic forms" (Austin and Leander, 2021, p. 92).

Collective memory shapes cultural identity, global politics, and an attitude in life towards temporal and spatial multiplicities. Memorializing differently situated positionalities draws on a community of divergent angles of vision. The autofictional, contrary to its common association with the narcissistic, offers the potential of the flow between fact and fiction to remember and imagine social and political solidarities that bears the worsening.

5 ACKNOWLEDGMENTS

I wish to thank the organizers of the Weizenbaum Conference for their invitation. I am grateful to my colleagues in the Security Vision research project, especially to Francesco Ragazzi, Ildikó Plájás, Ruben van de Ven, and Elka Smith for their generous feedback and support throughout. Many thanks to Mijke van der Drift, Lizzie Malcolm, and Daniel Powers for their guidance in developing and producing *Dear Chaemin*. This work was supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (SECURITY VISION, Grant Agreement No. 866535).

6 REFERENCES

1. Andrejevic, Mark. 2004. “The Work of Watching One Another: Lateral Surveillance, Risk, and Governance.” *Surveillance & Society* 2 (4). https://doi.org/10.24908/ss.v2i4.3359.
2. Austin, Jonathan Luke, and Anna Leander. 2021. “Designing-With/In World Politics: Manifestos for an International Political Design.” *Political Anthropological Research on International Social Sciences* 2 (1): 83–154. https://doi.org/10.1163/25903276-bja10020.
3. Balsom, Erika. 2018. “‘There Is No Such Thing as Documentary’: An Interview with Trinh T. Minh-Ha.” *Frieze*, November 1, 2018. https://www.frieze.com/article/there-no-such-thing-documentary-interview-trinh-t-minh-ha.
4. Chen, Nancy N. 1992. “‘Speaking Nearby:’ A Conversation with Trinh T. Minh–Ha.” *Visual Anthropology Review* 8 (1): 82–91. https://doi.org/10.1525/var.1992.8.1.82.
5. Corbett, Kevin J. 2016. “Beyond Po-Mo: The ‘Auto-Fiction’ Documentary.” *Journal of Popular Film and Television* 44 (1): 51–59. https://doi.org/10.1080/01956051.2015.1075954.
6. Di Salvo, Philip. 2021. “Solutionism, Surveillance, Borders, and Infrastructures in the ‘Datafied Pandemic.’” In *COVID-19 from the Margins: Pandemic Invisibilities, Policies and Resistance in the Datafied Society*, edited by Stefania Milan, Emiliano Treré, and Silvia Masiero, 164–70.
7. Dienderen, An van. 2017. “On *Lili* : Questioning China Girls through Practice-Based Research.” *Critical Arts* 31 (2): 72–86. https://doi.org/10.1080/02560046.2017.1357130.
8. Doubrovsky, Serge. 2013. “Autofiction.” *Auto/Fiction* 1 (1): 1–3.
9. Eng, David L., Jack Halberstam, and José Esteban Muñoz. 2005. “Introduction: What’s Queer About Queer Studies Now?” *Social Text* 23 (3-4): 1–17. https://doi.org/10.1215/01642472-23-3-4_84-85-1.
10. Forné, Anna, and Patricia López-Gay. 2022. “Autofiction and Film: Archival Practices in Post-Millennial Documentary Cinema in Argentina and Spain.” In *The Autofictional: Approaches, Affordances, Forms*, edited by Alexandra Effe and Hannie Lawlor. Palgrave Studies in Life Writing Series. London: Palgrave.
11. French, Martin, and Torin Monahan. 2020. “Dis-Ease Surveillance: How Might Surveillance Studies Address COVID-19?” *Surveillance & Society* 18 (1): 1–11. https://doi.org/10.24908/ss.v18i1.13985.
12. Haraway, Donna. 1988. “Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective.” *Feminist Studies* 14 (3).
13. ———. 2016. *Staying with the Trouble: Making Kin in the Chthulucene.* Experimental futures. Durham: Duke University Press.
14. Hong, Seong-Tae (홍성태). 2007. “Jumindeungnokjedowa Ilsangjeok Gamsisahoe: “Bakjeonghui Chegye” Reul Neomeoseo (주민등록제도와 일상적 감시사회: "박정희 체계"를 넘어서) [The Resident Registration System and Everyday Surveillance Society: Beyond the Bak Junghee System].” *Minjusahoewa Jeongchaegyeongu (민주사회와 정책연구) [Democratic Society and Policy Studies]* 12 (0): 321–45.
15. Jeong, Yeonbo. 2013. “Rethinking ‘Situated Knowledges’ Beyond Relativism: From the Fragmented Partiality to the Interconnected Partiality.” *Korean Feminist Philosophy* 19 (May): 59–83.

16. Lee, Hyemin, Horim Yi, G. Nic Rider, Don Operario, Sungsub Choo, Ranyeong Kim, Yun-Jung Eom, and Seung-Sup Kim. 2021. “Transgender Adults’ Public Bathroom-Related Stressors and Their Association with Depressive Symptoms: A Nationwide Cross-Sectional Study in South Korea.” *LGBT Health* 8 (7): 486–93. https://doi.org/10.1089/lgbt.2021.0007.
17. Lyotard, Jean-François. 1984. *The Postmodern Condition: A Report on Knowledge*. Theory and History of Literature, v. 10. Minneapolis: University of Minnesota Press.
18. Jun, Hye-Eun (전혜은). 2021. *Kwieo Iron Sanchaekagi (퀴어 이론 산책하기) [A Stroll Along/with Queer Theory]*. Yeoiyeon Iron 35. Seoul: Yeoiyeon (여이연) [Center for Women's Culture and Theory].
19. Park, Jun Hong, Ji-hye Baek, Jina Lee, and Heesun Chung. 2021. “Examining Socio-Spatial Exclusion of Minorities through the Case of Itaewon Coronavirus Outbreak.” *The Geographical Journal of Korea* 55 (2): 137–54.
20. Puar, Jasbir K. 2007. *Terrorist Assemblages: Homonationalism in Queer Times*. Next Wave. Durham: Duke University Press.Ramraj, Victor V. 2021. *COVID-19 in Asia: Law and Policy Contexts.* New York: Oxford University Press.
21. Ruin (루인). 2019. *Deo Naeun Miraeraneun Chakgak (더 나은 미래라는 착각) [The Illusion of a Better Future]. The Pinch (핀치).* Accessed September 10, 2020. https://thepin.ch/think/msb1N/ruin-column-2019-01.
22. Trinh, T. Minh-ha. 1990. “Documentary Is/Not a Name.” *October* 52: 76. https://doi.org/10.2307/778886.

Proceedings of the Weizenbaum Conference 2022:
Practicing Sovereignty. Interventions for Open Digital Futures

DIGITAL TWINS IN HEALTHCARE FOR CITIZENS

De Maeyer, Christel
Future Everyday, Department of Industrial Design, Eindhoven University of Technology, Eindhoven, Netherlands
c.a.a.d.maeyer@tue.nl

Lee, Minha
Future Everyday, Department of Industrial Design, Eindhoven University of Technology Eindhoven, Netherlands
m.lee@tue.nl

KEYWORDS

digital twins; digital healthcare; self-surveillance; surveillance; ethics

DOI: 10.34669/wi.cp/4.12

ABSTRACT

Digital twins are gaining attention in healthcare, especially in fields like hospital management, simulating surgeries, or providing personalized health. As digital replicas based on users' data, digital twins can inform citizens in-depth about their lifestyle, medical data, and biomedical data. Hence, there is the assumption that digital twins could facilitate preventative healthcare at home, bringing healthcare closer to citizens, yet there are underexamined ethical concerns. In this paper, we explore the ethics of digital twins based on citizens' perspectives on digital twins in healthcare via recent literature and research. Although digital twins have great potential, citizens have concerns about surveillance, data ownership, data accuracy, and personal and collective agency.

1 INTRODUCTION

We generate health data via a variety of applications—either through mobile applications or internet of things (IoT) devices. On top of this, biomedical data are increasingly digitized in hospitals and the practice of medicine in general is transitioning to the digital world. Against this backdrop, the notion of the digital twin in healthcare is emerging. A digital twin (DT) refers to a digital replica or a virtual presentation of a physical asset that serves as a digital counterpart (Grieves, 2014). This definition was formulated by Grieves in 2002, with the idea that a "digital information construct" connected to a physical object or asset, could be an entity on its own, becoming a "twin" of a person, object, or process by holding information about the original entity. Huang, (2022) refined the definition as it pertains to healthcare: "a digital twin for personalized health care service is a data-driven, interactive computerized model that aims to offer health-related information that properly simulates or predicts the health conditions of a particular person" (Huang, 2022, p. 12). This definition might not yet be complete and may be open for debate.

We see a digital twin as a layered technology that holds different data layers of a person, such as their environment, lifestyle, biomedical data, and other facets. A person's digital twin (DT) could also hold social health determinants[28]—that is, information on where citizens are born, live, and age, which are nonmedical factors that influence health. As a digital twin can predict, describe, and prescribe, the fact that a DT takes social health determinants into consideration for personalized medicine and treatments might give different insights into the backgrounds and needs of citizens in personalized healthcare. The notion that a DT could be available at home for citizens to manage, simulate, or predict their health has not been researched "in the field" thus far. Thus, a recent study (De Maeyer, 2022) that we conducted on DTs in healthcare at home offered insights on how citizens view and may use a DT. We found out that people would prefer DTs as qualitative representations rather than quantitative representations and wanted to be able to use both options in case of emergencies. Notably, the predictive features of DTs were not favored by citizens, mostly because they did not want to know what the future holds and wanted to live in the present. So, contrary to extant research (e.g., Huang, 2022) we may have to rethink definitions of DTs in light of what citizens want from the future of digital health. In this paper we propose a critical look at the ethics of DTs, such as how surveillance could lead towards new business models and policies in healthcare. We emphasize that how DTs are now conceptualized by professionals does not match citizens' expectations or needs, particularly since citizens we interviewed largely do not want their health states to be predictively portrayed by DTs.

[28] https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1

2 ETHICAL CHALLENGES

2.1 THE PUBLIC GOOD

We start with the presumed public good of DTs before diving into critiques. As DTs generate summative and predictive data on populations, they offer an opportunity to evaluate and create new insights that are broader than those offered interventionist healthcare. DT could be made available for academic research in different fields related to healthcare and well-being in general for preventative healthcare at home (Rasheed, 2020; De Maeyer, 2021). Citizens could give consent that their DT could be used for the greater good of a population. However, broad access is needed to get data sets that represent different layers of society and avoid discriminatory biases. Experts see a role for the government in overseeing the regulation of DTs, together with an expert board of different stakeholders, medical professionals, lawyers, ethicists, etc. Such an expert board could establish guidelines around this new technology. (De Maeyer, 2021; Boulos, 2021; Rasheed, 2020). In addition, an educational framework for educating professionals and citizens is an important aspect of embracing this emergent technology, starting with teaching children early on together with parents to create awareness on preventive healthcare (Barricelli, 2019; Rasheed, 2020). Yet the utopian vision of how healthcare DTs can serve the public good in a preventative manner is contentious, as we addressed below.

2.2 SELF-SURVEILLANCE

Simply put, self-surveillance means paying attention to one's own behavior. In 2007, Kevin Kelly and Gary Wolf popularized the quantified-self movement. Early self-tracking devices and mobile apps became available on the market and could be used to track different bodily aspects, such as physical activity, mood, calories, sleep and so forth. This is one of the drivers of the idea of DT, together with other digitized health data that are available in hospitals and with medical professionals. From a sociological perspective, critics have expressed the notion that the quantified self could empower individuals to manage their own health, going from "'health is the responsibility of my medical professionals surrounding me' to 'I'm responsible for my health'" (Swan, 2012, p. 108). Much of this discourse still holds for a DT concept, especially if it were be available in a home environment. These self-tracking apps or devices could be imposed or pushed on citizens by different stakeholders for different purposes—for instance, to get personal information for a given person, which we widely saw with COVID-19 tracking applications. It can help when an individual consents to tracking their heartrate or blood pressure, as well as their use of medication—this may offer insights that are useful before and after a surgery. This is regularly done when agreed between patient and

GP, for instance. Yet, the thinking is that such personal information would then be available in the DT of that specific citizen, allowing near real-time monitoring.

As health applications are easily available today, they become part of our daily lives, and self-surveillance has almost become a norm in our society—it is now nearly an obligation to actively observe oneself (Han, 2017). Doctors are quoted as saying "within 10 years I want to be able to open my laptop during consultations to view the stress data of the patient sitting in front of me"[29]. This illustrates what self-tracking modes may entail in the upcoming years for healthcare in general. Lupton (2014) defines five modes of self-tracking. There is private self-tracking, referring to voluntary self-tracking activities, and pushed self-tracking, referring to self-tracking coming from another agent or actor, and usually encouraged externally by a general practitioner, for example. Communal self-tracking involves voluntary sharing of personal data in communities, e.g. sharing physical activity data in Strava.[30] Hence, imposed self-tracking, usually by other parties, can be expected in health care environments, but also in work environment to optimize citizens' labor in general; this may easily become exploitative self-tracking, where self-tracked personal data are repurposed for other means, usually commercially, such as for reward systems as customer' loyalty programs (Lupton, 2014). These modes also could apply to a DT as it can push, impose, or exploit citizens. What is different about DTs is that their status as *replicas* of citizens while taking a *predictive* stance; by predicting people's futures as replicas, DTs can push, impose, or exploit citizens to change their behavior via *predicted future states*. Rather than intervening on current health states or ailments, citizens are exposed to and can be expected to act based on data-driven future versions of themselves. Hence, self-surveillance in the present paves the way for forecasted future surveillance of citizens, further endangering our agency and privacy.

2.3 AGENCY

Agency is discussed in two ways. One is the loss of collective agency, and the other is personal agency, and the two of these are related. Collective agency, in the context of this paper, refers to the democratization of healthcare in which, as per above (the public good), we can exercise preventative healthcare through DTs, in which society at large benefits through data-sharing and preventative health management. Researchers have stated that DTs could be social equalizers but they could also broaden the digital divide gap; DTs could lead to social sorting, social segmentation, and increasing discrimination (Bruynseels, 2018; Boulos, 2021). The idea that DTs would enhance humans could also lead towards a new class of people, disrupting democratic processes when citizens are treated

[29] https://www.tijd.be/dossiers/de-meetbare-mens/burn-outs-voorkomen-met-data/10351299.html
[30] https://www.strava.com/

differently and unfairly through DTs (Fukuyama, 2002). With broad access to DTs for citizens, social sorting or specific segmentation is a worry.

In relation to collective agency, we discussed the loss of personal agency perceived by the participants we interviewed (De Maeyer, 2022). As several participants noted, a DT could be connected to health insurance providers, which could then see how a citizen performs and adjust insurance pricing accordingly. Furthermore, insurance providers could create reward systems, according to an expert: "*incentives for sharing data could be rewarded through vouchers or loyalty cards"* (De Maeyer, 2021). Looking at business models in this perspective, one of the experts also mentioned, *"the danger of connecting financial information to a digital twin might evolve towards more of an economic exchange system than a healthcare system"* (De Maeyer, 2021*).* There is a clear divide between what experts think and citizens think. Citizens argue that they do not want their DTs to be connected to insurance providers or financial information. One citizen stated: *"I think it should be protected, if my hospitalization insurance is giving up on me, because according to them I don't fall within the standards, I don't want that, so I don't want them to know, actually"* (De Maeyer, 2022*).*

Other participants commented further that they would prefer an offline system in which they could control and synchronize their data when they saw fit. In other words, people want to have control of the DT and its data. Due to the close link between the digital replica and the citizen, the question arises of whether people will be able to make the right decisions autonomously and whether they are able to interpretate the data correctly. Furthermore, the proposed decisions DTs make are likely to be algorithmic. This may be a new form of "dataism," in which a DT becomes a "medical patronizing system" (Bruynseels, 2018). A human should be in the loop, not only to support decision making but also to check the results presented by a DT (Rasheed, 2020). Yet this may not be enough considering privacy issues.

2.4 PRIVACY IN AND OF DIGITAL TWINS

Barricelli (2019) explains that deploying DTs would demand seamless connections, sensors, and know-how to foster interest in DTs for researchers and doctors but also for the citizens. As a DT makes use of cloud-based services to collect health data, the privacy and robustness of this technology is of major importance, especially due to the medical and lifestyle information that a DT holds. The EU's General Data Protection Regulation (GDPR), which has been in force since April 25, 2018,[31] is a step forward towards protecting individuals' privacy (Rasheed, 2020). A problem is that the GDPR

[31] https://gdpr-info.eu/

is broad and not adequate for use with DTs. Another issue is how DTs designed and developed outside the EU may or may not be compliant with GDPR. As with other digital applications, it is unclear how citizens become aware of the aggregation of DT data from within Europe with non-EU compliant data and where the data handling responsibility resides. What makes DTs different from other digital applications is that they are taken to be the "replica" of citizens; if surgery simulations are undertaken or health predictions made with DTs, there are additional privacy concerns. For instance, a DT could be hacked or be infected with viruses, meaning that people are greatly vulnerable due to the sensitive health data and predictions in their DTs. If a citizen's DT is hacked, then inaccurate health forecasting could be implanted in the DT, which could impact high-stake situations like surgeries that depend on data held in the DT. Beyond concerns about healthcare insurance premiums, real-time life or death decisions—e.g., through wrongly simulated operations—become a major concern when privacy cannot be guaranteed.

2.5 HEALTH FORECASTING AND SIMULATIONS

A strong reason not to predict people's health through DTs is that citizens may not want this. While forecasting and simulations are one of the features of a DT, we noticed that citizens, with the exception of one outlier, were not keen on using that feature. The notion of forecasting health felt too confrontational, together with the excessive number of uncertainties and variables that influence our wellbeing, like a user argued: "*for me, personally that is scary, I feel more vulnerable than before, friends that are dying, it all becomes so visible"* (De Maeyer, 2022).

As Braun (2021) puts it, if a health prediction points to a severe illness, it will change the life of the citizen or patient; the DT will influence thinking and might have power over the person, limiting their freedom. This relates to earlier discussions on surveillance: Influencing people's presents and futures via simulated future health states can severely limits their collective and personal agency. Yet, interviewees welcomed the prospective use of DTs as reflective tools rather than predictive replicas. Citizens can take a reflective stance on what being healthy may individually mean, according to our study. Thus, participants preferred a qualitative representation rather than a quantitative representation. For one, the interpretation of quantitative data would be hard for some to understand and to cope with. A qualitative representation, like a digital painting as a landscape of one's moods, could represent their well-being while leaving room for personal interpretation. But, quantitative representations—e.g., calories consumed as graphs—could create and enhance feelings of vulnerability. In sum, citizens and experts may have differing opinions on how they expect DTs to develop.

3 CONCLUSION

In this paper, we explored citizens' perspectives from a previous study (De Maeyer, 2022) and interwove these explorations with background research on digital twins in healthcare. We covered the surveillance aspects of a DT in healthcare from the citizens' perspective, building on the modes of self-tracking practices. DTs are said to offer public good in evaluating and analyzing the mass of data that will become available on a population, if citizens consent, thus democratizing healthcare. But we see a divide in the views of experts and citizens, mainly in the need for DTs to have forecasting abilities. People would prefer a DT that served more as a tool for reflection than forecasting. Data protection, privacy, and the robustness of the technology should be ensured, but such practices still leave out deeper ethical concerns, such as the surveillance of currently healthy "sick people of the future," thus endangering our collective and personal agency.

4 REFERENCES

1. Grieves, M. (2014, January 3). Digital Twin: Manufacturing Excellence through Virtual Factory Replication. USA.
2. Huang, P. (2022). Ethical Issues of Digital Twins for Personalized Health Care Service: Preliminary Mapping Study. *Journal of Medical Internet Research, 24*(1), e33081.
3. Barricelli, B. R. (2019). A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications. *IEEE Xplore, 7*, pp. 167653-167671.
4. Rasheed, A. (2020). Digital Twin: Values, Challenges and Enablers From a Modeling Perspective. *IEEE Access*, 21980-22012.
5. Swan, M. (2012, Sep 12). Health 2050: The Realization of Personalized Medicine through Crowdsourcing, the Quantified Self, and the Participatory Biocitizen. *Journal of Personlized Medicine, 2*(3), 93-118.
6. Han, B.-C. (2017). Psychopolitics: Neoliberalism and New Technologies of Power. Verso.
7. Lupton, D. (2014). Self-Tracking Modes: Reflexive Self-Monitoring and Data Practices. *SSRN Electronic Journal*.
8. Bruynseels, K. (2018). Digital Twins in Health Care: Ethical Implications of an Emerging Engineering Paradigm. *Frontiers in Genetics*, 31.
9. Boulos, K. (2021, July). Digital Twins: From Personalised Medicine to Precision Public Health. *Journal of Personalised Medicine, 11*(745).
10. Fukuyama, F. (2002). *Our Posthuman Future: Consequences of the Biotechnology Revolution.* Published by Farrar, Straus and Giroux. Macmillan.
11. Braun, M. (2021). Represent me: please! Towards an ethics of digital twins in medicine. *Journal of Medical Ethics;, 47*(6), 394-400.
12. De Maeyer, C. (2021). Future outlook on the materialisation, expectations and implementation of Digital Twins in healthcare. https://doi.org/ 10.14236/ewic/HCI2021.18. *34th British HCI Conference (HCI2021)* (pp. 180-191). London: BCS Learning & Development Ltd. Proceedings of the BCS 34th British HCI Conference 2021, UK.
13. De Maeyer, C. (2022). I feel you. *HCSE* (paper in publication process). Springer-LNCS series.

Proceedings of the Weizenbaum Conference 2022:
Practicing Sovereignty. Interventions for Open Digital Futures

DRAWING AS A FACILITATOR OF CRITICAL DATA DISCOURSE

REFLECTING ON PROBLEMS WITH DIGITAL HEALTH DATA THROUGH EXPRESSIVE VISUALIZATIONS OF THE UNSEEN BODY LANDSCAPE

Kuksenok, Kit
Independent
Berlin, Germany
ksenok@protonmail.com

De Maeyer, Christel
Future Everyday, Department of Industrial Design, Eindhoven University of Technology
Eindhoven, Netherlands
c.a.a.d.maeyer@tue.nl

Lee, Minha
Future Everyday, Department of Industrial Design, Eindhoven University of Technology
Eindhoven, Netherlands
m.lee@tue.nl

KEYWORDS

digital health data; self-tracking; data visualization; data ethics

DOI: 10.34669/wi.cp/4.13

ABSTRACT

In a 1.5-hour workshop, we used drawing and self-reflection prompts to facilitate a value-driven discussion of personal and institutional data practices. Activities included mark-making in time with one's heartbeat, creating an inventory of one's personal data, and creating a qualitative personal health visualization. This article details the workshop structure and exercises and includes a summary of the discussion, which constructively encompassed both the empowering and the uncomfortable aspects of digital health data collection in a constructive manner. The workshop's design used the format of hands-on, expressive drawing activities to enable participants to achieve depth and breadth in a relatively short discussion about personal health, data autonomy, institutional trust, and consent. Critical discourse about data, especially health data, is a valuable experience for every person whose health data has been or is being collected; and approaches that take personal data as a starting point can support the practice of digital/data sovereignty more broadly.

1 INTRODUCTION

This article describes and reflects on a workshop about the challenges of digital health data. We first review the background on visualizing the body interior, which established common ground; then we describe the drawing exercises and self-reflection prompts used; lastly, we share and reflect on the key themes which emerged from the discussion.

Data that is digitally gathered, such as steps walked or hours slept, provides a quantified summary of behavior that can be difficult for the individual to interpret or to act upon. Data collected by individuals can also be aggregated and (mis)used in unexpected ways. Quantitative data are thought to be the backbone of predicting future health, but they can hamper rather than support everyday citizens' understandings of their health. Predictions based on quantification provide a narrow look at what it means to be healthy or well, and proprietary technologies can have changing, unverifiable, and systematically biased inaccuracies. One motivation for self-tracking through "quantified-self" interventions is to render an unseen body experience visible and to control some aspect of life, but the available tools may not only fail to deliver on many expectations of visibility and control but also introduce new sources of obscurity and powerlessness (Kuksenok & Satsia, 2021), both at a bodily and a societal/institutional level.

The 1.5-hour workshop we hosted combined expressive drawing exercises and discussion prompts to highlight problems with digital health data. We focused on rich data generation, including qualitative data and qualitative representations of quantitative data. Qualitative data, such as visuals or text, resist common summarization practices that quantitative data affords and are thus rich sites for considering the role of algorithmic classification and summary visualizations in reducing and containing data. Qualitative representations of data can also offer insights into individuals' "lived experience," as people tend to express themselves more metaphorically, i.e., "I feel stuffed" or "I slept like a baby" (Lockton et al., 2017). We based our work on prior workshops (Kuksenok, 2022) that featured daily artistic practices of re-thinking what data can be situated within data feminism (D'Ignazio & Klein, 2020) as a framework for reflection on, and critical discourse on, personal and institutional data practices.

2 BACKGROUND

What does it mean to use data to visualize and understand the interior unseen body landscape? To help established a common starting point, the workshop began with a brief round of participant introductions and a short review of several ideas from existing literature on health data from different fields. The dozen participants in the workshop had different professional experiences, but all had

some personal interest in digital health data. The selected anchoring references, provided as a single-page handout in the workshop and summarized below, situate the expressive exercises (described in Section 3) in relation to the theme of the conference: practicing [digital/data] sovereignty.

Data collected through and about the body can be generative and insightful creative material, "like paint or paper, offering a new way of seeing and engaging with the world" (Lupi & Posavec, 2018). However, data used to render the body more visible—as a "screen body"—may sometimes lead individuals to mistrust their own senses and assume manageability. As the following quote illustrates:

> The visual image of the data [contemporary technologies of measuring and observing the body] generate are often privileged as more "objective" than the signs offered by the "real", fleshy body and the patients' own accounts of their bodies. … As part of the project of seeking security and stability, such technologies attempt to penetrate the dark interior of the body and to render it visible, knowable and thereby (it is assumed) manageable. (Lupton, 2016, p. 53, citing others)

Motivations for self-tracking include not only the desire to observe but to gain control: (1) reducing or eliminating uncertainty, (2) truthfully observing a bodily experience, an (3) directing behavior change. However, methods for self-tracking entail losing control, such as when: (2) new sources of uncertainty are encountered, (2) "objective" data brings disconnection from the subjective experience, and (3) behaviors are influenced in unintended ways (Kuksenok & Satstia, 2021). A relative loss of control not only includes the immediate and behavioral but also subtle aspects of data's role in society. For example:

> [Although it can be argued] that self-tracking is an alternative data practice that is a form of soft resistance to algorithmic authority and to the harvesting of individuals' personal data. They argue that self-tracking is... "a profoundly different way of knowing what data is, why it is important, who gets to interpret it, and to what ends. … However, the issue of gaining access to one's data remains crucial to questions of data control and use. While a small minority of technically proficient self-trackers are able to devise their own digital technologies for self-tracking and thus exert full control over their personal information, the vast majority must rely on the commercialized products that are available and therefore lose control over where their data are stored and who is able to gain access." (Lupton, 2016, p. 133; citing Nafus & Sherman, 2014).

Data, especially personal health data, has the capacity for betrayal (ibid.) because it can be used by institutions as mechanisms of surveillance and control. One example of this is when employers require workers to report measures of health and uses these to inform health insurance contribution (O'Neil, 2016). In response to this, contemporary artists have explored the possibility of adapting existing body-observation and body-measurement tools for counter-normative goals (Kuksenok &

Satsia, 2021; Satsia & Kuksenok, 2021). Such subversive body projects may shift the emphasis away from "self-knowledge through numbers" toward "[treating] digital self-tracking devices not as means of self-discovery but as tools for inventing oneself as something new and not yet imagined"; instead of "body projects" that "define progress, success, and satisfaction in terms of the exterior form of the body … [toward a] counter-normative and more liberating digital body project would perhaps be purposefully goal-unoriented"; and instead of "game design elements" which in practice "do not make self-tracking endeavors truly fun, playful, or pleasurable," "focus on the quality of one's interior experience... thereby adopting a counter-normative way of experiencing the body and evaluating how one feels" (Sanders 2017, pp. 21–22). Lastly, we shared the list of data feminism principles (D'Ignazio & Klein, 2020), which stress reflection on context and the examination power dynamics built into data objects, as a starting point for articulating one's values about personal and institutional data practices.

3 EXERCISES & PROMPTS

Following introductions and context-setting, we went through a series of three exercises that combined drawing and self-reflection. The materials provided were color pens and markers, graph paper, and tracing paper. The handouts with references (as summarized in Section 2) also included one-sentence summaries of the three exercises and the prompts (below, these prompts are *italicized*).

Resonant heartbeats. *Take 1.5 minutes to make tick marks with a pen/pencil on a piece of paper every time your heart beats.* This exercise, adapted from (Lupi & Posavec, 2018), has three key goals within the context of the workshop: (1) It centers on the body, as it can be challenging to find a heartbeat; (2) It supports starting a discussion about data observation—when did you make the tick-mark? Did observing the heartbeat change it? (3) It creates a shared, embodied experience through sound.

Data inventory. List as many existing personal data sources as you can. For each: What would be an obvious finding from this data? What would be a surprising finding? For the whole list: What data sources are complementary? What data sources help validate a surprising finding? When moving on to the next exercise, the participants are encouraged to keep adding to this list if new data sources come to mind. As in prior workshops, some are surprised by how long this list can become.

Data archaeology and re-activation. Decide on a personal topic to retroactively explore for some time interval in the past (the longer the better; months or even years), ideally something that is still relevant today (widely applicable examples include sleep, mood, movement, or food). Start with a memory of key event dates, listing them; then create an accordion with the tracing paper (this was

demonstrated) with as many folds as there are events, drafting the first major timeline. Within this, fill in the middle bits, using, for example, a calendar or other sources of data. Tracing paper can be layered to make revisions and additional notes or participants can use one sheet of paper for one type of data (e.g., sleep) and a second for another (e.g., specific test results). Within the workshop, participants are free to leave placeholders or "coded" notes to self to maintain privacy. Further prompts:

- Were you actively tracking anything during this timeframe? Related or unrelated to the chosen subject? Or passively tracking?
- Are data tracked and stored but inaccessible to you? What or how can that data be retrieved, and would it be useful?
- Does anything emerge as an area of interest—something that maybe you would want to look more closely at?
- How did you deal with missing data or uncertain data? Or data from multiple sources?
- What has been the role of tracking and reflecting on data for you so far? Short versus long-term data tracking? Do any new possibilities arise?

Within the drawing exercises, the orientation toward one's health data (supplemented by other, non-health-specific sources) is generally approached from the perspective of possibility and of exploring the potential benefits of long-term self-reflection through data in a way that directly corresponds to participants' own interests. After a short break, we built on this shared experience of embodied self-reflection within a more abstract, value-oriented discussion.

4 DISCUSSION

In this section, we summarized some of the key themes that arose during the discussion among the dozen participants. Although the exercises initially centered on individual experience, both the follow-up prompts and the context of the conference ("practicing sovereignty") contributed to themes arising about institutional trust.

How is health data tracked? The discussion distinguished broadly between **passive or active** methods, based on whether any action was needed to record an event. This distinction has ethical, epistemic, and usability implications. At the level of usability and user experience (UX), the discussion brought up the difficulty of active tracking. Food intake tracking, for example, is typically a manual activity, where US-based apps use US-based nutrition and product databases, making it more difficult to use in a non-US context. Self-tracking activities can be more sustained when there is a need; this was the subjective experience of participants who tried self-tracking, and it is generally observed in, for example, research on self-tracking for the management of diabetes, where individuals

need to monitor metabolic state information and well as medical guidance. Most consumer food tracking applications, even when driven by a need, have many opportunities for UX friction, which degrades the accuracy of data: Are users weighing every bit of food they eat to track their food intake, and if a database is used to simplify this process, does it reflect the products in their region? There are accuracy challenges in any tracking application, but the epistemic challenge goes beyond that. Even if accuracy is well-understood, it is never 100%, and implied causal links may not be applicable. Within the realm of medical testing and screening, testing and observation must be justified, especially for tests associated with higher false-positive rates. Meanwhile, consumer tracking applications take the opposite approach, offering quick fixes, although the scale at which meaningful bodily change occurs is typically long. Lastly, on a data-ethical level, the group expressed uncertainty about how much data is being tracked passively on consumer phones; and a concern about the lack of awareness among the general population. Among the conference attendees, there was a high degree of awareness of the potential pitfalls of digital data, but even within the small group, there were different mental models and degrees of awareness of the capacity of health data to be misused.

In considering **autonomy and consent**, the group generally agreed that "everyone should own their own health data" and control it, which would mean any apps involved would not be free. One mechanism for data ownership has been local on-device storage for mobile tracking apps, such as for one menstruation app (Drip) that was mentioned. Following the ban on abortion in many US states, the privacy policies of period-tracking apps have come under scrutiny—the concern is that these apps maintain detailed history of fertility and sexual activity that could be used against the interests of their users. Although on-device storage is a useful mechanism in some cases, it would not necessarily prevent user from being legally compelled to share their data; furthermore, the complexity of each of these apps is such that the consumer must trust both the app and the app ecosystem to operate in good faith. In the discussion of consent, parallels were drawn to the EU's General Data Protection Regulation (GDPR), with participants noting that consent should be understood in relation to a specific purpose and that autonomy requires the capacity to accept the consequences of providing or not providing health data and knowing the consequences of either sharing or refusal.

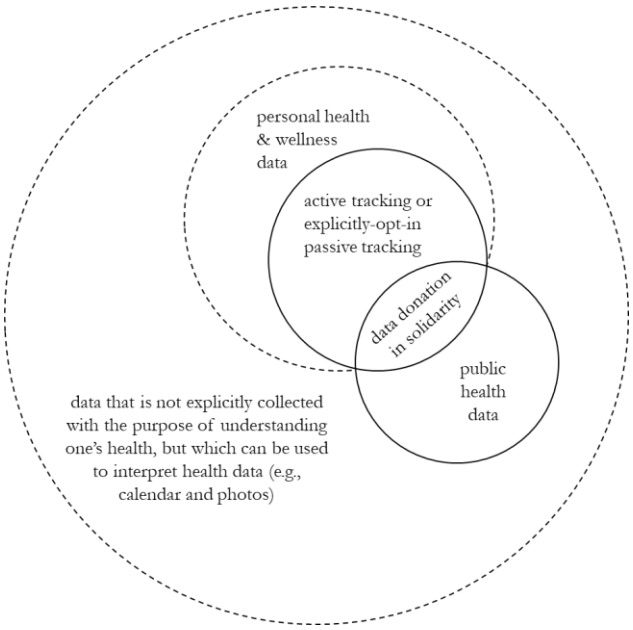


Figure 1. A What is health data? Several overlapping but distinct types of data came up in the discussions.

What is the purpose of collecting personal health data? Aside from general wellness-related goals or the context of managing one's health within a medical treatment plan, the group also recognized public goals. Personal and public health datasets can overlap when personal data can be donated to a trusted institution (despite some possible risk even when anonymity is maintained) as an act of **solidarity**. This is typically organized by institutions requiring active, opt-in consent. For example, female citizens in Belgium aged 50–69 can participate in a breast cancer screening examination, which can be aggregated at the population level. Such screenings are based on informed consent, which is explained on the screening websites, and derive their trustworthiness from being government initiatives intended to create a social good. Nevertheless, this aggregate data still has the potential to adversely affect particular populations if it informs policy connected to health insurance (depending, of course, on the context and content of such policy). The risks associated with any aggregation of personal data into public datasets, by either public or private institutions, depend on the specific vendors and technologies used. Both personal and aggregated/public health data are within a broader sphere that includes health data that users do not explicitly consent to (such as passive step tracking by a mobile device that a user has forgotten about), or non-health-specific data that could potentially be used together with the health data. This broader realm of data is not typically easily usable through tracking software, but it is subject to similar data-ethical issues of ownership and storage. Aggregate data, such as the data in public health datasets, contributes to another challenge that the discussion touched on: that medical professionals may not have the time to look deeply into the data available, or when they do, they may pay more attention to the aggregate than to the subjective experience.

This discussion had wide-ranging themes, which we have summarized above. Although the facilitators asked follow-up and clarification questions, the discussion prompts were either general and grounded in personal data reflection (such as those in Section 3) or open-ended and related to the anchoring references (such as those in section 2). The breadth and depth of this discussion reflects the variety of perspectives that the participants offered. The facilitator took care to keep the discussion constructive, by building connections between recurring themes—*When is a particular data practice worth the risk? What is the risk?*—instead of initiating a polarizing debate by asking questions on whether a particular data practice is good or bad. Participants had different professional and personal relationships to digital health and tracking applications, and even in the shared context of a conference on data sovereignty, these different backgrounds led to different perceptions and interpretations of the practical state of data tracking in relation to the shared ethical sense that autonomy and consent are essential. For example: although there was some consensus that even not entirely risk-free activities (active self-tracking for specific reasons; donations to a public health dataset) could be worth undertaking within the context of trust and credibility (trusting a credible app; trusting a credible institution), no mechanisms for establishing or recognizing this trust were suggested. All topics in this discussion are the subjects of active research, but they are also deeply relevant to everyday citizens, whose data (health and beyond) is collected, aggregated, and used extensively. Thus, we believe it was valuable to facilitate a workshop where these subjects could be explored actively (through drawing and discussion), rather than passively (through reading popular articles, which can be polarizing).

5 CONCLUSION

The goal of our workshop was to enable participants to (1) try out new ways of encountering tracked and health data for reflection, (2) practice applying critical and reflective data practices to health data and beyond, and (3) experience community data reflection in action and reflect on their values with respect to data. As documented in this article, these goals were addressed through drawing and discussion over the course of 1.5 hours. Although the conference context provided some shared background about digital technologies generally, the participants' backgrounds with respect to health data varied widely. The topics in the discussion are subjects of active research but are not typically the subject of casual public discourse. Even when they are brought up, it can be difficult to reflect both the empowering and the uncomfortable aspects of digital health data collection in a constructive manner. This example of this workshop illustrated how the format of hands-on, expressive drawing activities can lend depth and breadth even in a relatively short discussion among strangers. We held this workshop because we believe that critical discourse about data, especially health data, is a

valuable experience for every person whose health data has been or is being collected; and that starting with personal data can support crucial discourse on other aspects of how data is produced and handled, to practicing digital/data sovereignty more broadly.

6 ACKNOWLEDGMENTS

Kit Kuksenok's work on facilitating community data practices is partly supported by the School of Commons (ZHdK), in the READ (Research Ecologies and Archival Development) Lab. Prior teaching through with the School of Machines, Making, and Make-Believe and Ora Collective, as well as collaborations with artist and researcher Marisa Satsia and journalist Saga Briggs, were also essential to the development of workshop materials.

7 REFERENCES

1. D'Ignazio, C., Klein, L.F. (2020). *Data Feminism*. MIT Press.
2. Kuksenok, K., Satstia, M. (2021). Know thy Flesh: What Multi-disciplinary Contemporary Art Teaches Us about Building Body Knowledge. In Proceedings of xCoAx 2021.
3. Kuksenok, K. (2022). "Critical Data Practice at Home and with Friends." *Critical Coding Cookbook*. Retrieved August 28, 2022. https://criticalcode.recipes/contributions/critical-data-practice-at-home-and-with-friends
4. Lockton, D., Ricketts, D., Aditya Chowdhury, S., Lee, C. H. (2017). Exploring qualitative displays and interfaces. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (pp. 1844-1852).
5. Lupi, G., Posavec, S. (2018). *Observe, collect, draw!: A visual journal*. Princeton Architectural Press.
6. Lupton, D. (2016). *The Quantified Self.* John Wiley & Sons.
7. O'Neil, C. (2016). *Weapons of Math Destruction*. Broadway books.
8. Sanders, R. (2017). Self-tracking in the digital era: Biopower, patriarchy, and the new biometric body projects. *Body & Society*, 23(1), 36-63
9. Satsia, M., & Kuksenok, K. (2021). From Data to Matter: Anti-Systematic Interventions and Explorations of the (Micro) biopolitical Organism. *Proceedings of Polititcs of the machines-Rogue Research 2021 3*, 205-213.

Proceedings of the Weizenbaum Conference 2022:
Practicing Sovereignty. Interventions for Open Digital Futures

DEFENDING INFORMATIONAL SOVEREIGNTY BY DETECTING DEEPFAKES?

RISKS AND OPPORTUNITIES OF AN AI-BASED DETECTOR FOR DEEPFAKE-BASED DISINFORMATION AND ILLEGAL ACTIVITIES

Tahraoui, Milan
Berlin School of Economics and Law
Berlin, Germany
milan.tahraoui@hwr-berlin.de

Krätzer, Christian
Magdeburg University
Magdeburg, Germany
kraetzer@iti.cs.uni-magdeburg.de

Dittmann, Jana
Magdeburg University
Magdeburg, Germany
jana.dittmann@iti.cs.uni-magdeburg.de

KEYWORDS

deepfake detection; digital sovereignty; remote ID proofing.

DOI: 10.34669/wi.cp/4.14

ABSTRACT

This paper will first investigate possible contributions that an AI-based detector for deepfakes could make to the challenge of responding to disinformation as a threat to democracy. Second, this paper will also investigate the implications of such a tool—which was developed, among other reasons, for security purposes—for the emerging European discourse on digital sovereignty in a global environment. While disinformation is surely not a new topic, recent technological developments relating to AI-generated deepfakes have increased the manipulative potential of video and audio-based contents spread online, making it a specific but important current challenge in the global and interconnected information context.

1 INTRODUCTION

Google has recently forbidden the use of its Colaboratory (Colab) service—one of the most popular platforms online to train machine-learning and AI systems with free computational resources—to generate deepfakes.[32] This is one example among others of the increasing risks perceived to be associated with deepfakes. These risks have motivated public authorities, such as the European Commission,[33] Cyber Administration of China,[34] as well as global leading private firms, such as Google and Meta,[35] to regulate their generation and circulation. One of the most commonly perceived risks with deepfakes, beyond so-called "revenge porn" or harmful application cases, is the anticipated

---

[32] TechRadar.com, "Google is cracking down hard on deepfakes", 31 May 2022, at https://www.techradar.com/news/google-is-cracking-down-hard-on-deepfakes; reseach.google.com, "Colaboratory: Frequently Asked Questions", at https://research.google.com/colaboratory/faq.html (page visited on 28 August 2022): "We prohibit actions associated with bulk compute, actions that negatively impact others, as well as actions bypassing our policies. The following are disallowed from Colab runtimes: […] creating deepfakes."

[33] Reuters.com, "Exclusive: Google, Facebook, Twitter to tackle deepfakes or risk EU fines", 14 June 2022, at https://www.reuters.com/technolog[33] TechRadar.com, "Google is cracking down hard on deepfakes", 31 May 2022, at https://www.techradar.com/news/google-is-cracking-down-hard-on-deepfakes; reseach.google.com, "Colaboratory: Frequently Asked Questions", at https://research.google.com/colaboratory/faq.html (page visited on 28 August 2022): "We prohibit actions associated with bulk compute, actions that negatively impact others, as well as actions bypassing our policies. The following are disallowed from Colab runtimes: […] creating deepfakes."

[33] Reuters.com, "Exclusive: Google, Facebook, Twitter to tackle deepfakes or risk EU fines", 14 June 2022, at https://www.reuters.com/technology/google-facebook-twitter-will-have-tackle-deepfakes-or-risk-eu-fines-sources-2022-06-13/.

[33] Reuters.com, "China issues draft rules for fake in cyberspace", 28 January 2022, at https://www.reuters.com/world/china/china-regulator-issues-draft-rules-cyberspace-content-providers-2022-01-28/.

[33] Meta, "Enforcing Against Manipulated Media", 6 January 2020, at https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/.

[33] See for instance, Markus Appel, Fabian Prietzel, "The detection of political deepfakes", *Journal of Computer-Mediated Communication*, 2022, Vol. 27, No. 4, at https://academic.oup.com/jcmc/article/27/4/zmac008/6650406; Matthew Bodi, "The First Amendment Implications of Regulating Political Deepfakes", *Rutgers Computer and Technology Law Journal*, 2021, Vol. 47, No. 1, pp. 143-172; Marc Jonathan Blitz, "Deepfakes and Other Non-Testimonial Falsehoods: When is Belief Manipulation (Not) First Amendment Speech?", *Yale Journal of Law& Technology*, 2020, Vol. 23, No. 3, pp. 160-300; Bobby Chesney and Danielle Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security", *California Law Review*, 2019, Vol. 107, No. 6, pp. 1753-1820.

[33] See for instance, Ido Kivovaty, "The international law of cyber intervention" *in* Nicolas Tsagourias and Russel Buchan (eds.), *Research Handbook on International Law and Cyberspace*, 2021, Cheltenham (UK)/Northampton (US)Edgar Elgar Publishing, 2nd ed., 2021, xxviii-634p., pp. 97-112, p. 104; Nicholas Tsagourias, "Electoral Cyber Interference, Self-Determination and the Principle of Non-Intervention in Cyberspace" *in* Dennis Broeders and Bibi van den Berg (eds.), *Governing Cyberspace: Behavior, Power, and Diplomacy*, Lanham/Boulder/New York/London, 2020, Rowman & Littlefield, vii-327p., pp. 45-63.

[33] This contribution is based on the research project *FAKE-ID: Videoanalyse mit Hilfe künstlicher Intelligenz zur Detektion von falschen und manipulierten Identitäten (meaning "AI-based video analysis to detect false and manipulated identities")*, financed by the German Federal Ministry of Education and Research (BMBF) within the framework of the research programme *Künstliche Intelligenz in der zivilen Sicherheitsforschung ("AI in civil security research")* (FKZ: HWR/FÖPS 13N15737, OVGU 13N15736).
https://www.reuters.com/technology/google-facebook-twitter-will-have-tackle-deepfakes-or-risk-eu-fines-sources-2022-06-13/.

[34] Reuters.com, "China issues draft rules for fake in cyberspace", 28 January 2022, at https://www.reuters.com/world/china/china-regulator-issues-draft-rules-cyberspace-content-providers-2022-01-28/.

[35] Meta, "Enforcing Against Manipulated Media", 6 January 2020, at https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/.

facilitation and intensification of disinformation spreading and other forms of manipulation online.[36] Against this backdrop, this paper will first investigate a number of possible contributions that an AI-based detector for deepfakes could make to the challenge of responding to disinformation as a threat to democracy.[37] Additionally, this paper seeks to analyze and frame the implications of such a tool within the emerging European discourse on digital sovereignty in a global environment.[38] While disinformation is certainly not a new topic, recent technological developments relating to artificial intelligence ("AI")-generated deepfakes have increased the manipulative potential of video and audio-based contents available online, making it a specific, important current challenge in the global and interconnected information context.

One important contextual background element is the current global competition for leadership taking place in the field of artificial intelligence and machine-learning technologies. This primarily involves the two global leading technological poles—namely the United States and China—but also the European Union as well as other states such as Russia. This competition also intervenes in the AI/machine-learning ("ML") regulatory field, with the European Union and China being at the forefront of drafting non-sectorial AI regulatory frameworks.[39] Yet, there is currently no global consensus on AI/ML regulation, despite some important developments such as the adoption of the 2021 UNESCO recommendation on the ethics of artificial intelligence.[40]

In this context, with the European Commission's Proposal for an AI regulation, the European Union is currently trying to occupy this space to establish itself as a global standard-setter with a

---

[36] See for instance, Markus Appel, Fabian Prietzel, "The detection of political deepfakes", *Journal of Computer-Mediated Communication*, 2022, Vol. 27, No. 4, at https://academic.oup.com/jcmc/article/27/4/zmac008/6650406; Matthew Bodi, "The First Amendment Implications of Regulating Political Deepfakes", *Rutgers Computer and Technology Law Journal*, 2021, Vol. 47, No. 1, pp. 143-172; Marc Jonathan Blitz, "Deepfakes and Other Non-Testimonial Falsehoods: When is Belief Manipulation (Not) First Amendment Speech?", *Yale Journal of Law& Technology*, 2020, Vol. 23, No. 3, pp. 160-300; Bobby Chesney and Danielle Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security", *California Law Review*, 2019, Vol. 107, No. 6, pp. 1753-1820.

[37] See for instance, Ido Kivovaty, "The international law of cyber intervention" *in* Nicolas Tsagourias and Russel Buchan (eds.), *Research Handbook on International Law and Cyberspace*, 2021, Cheltenham (UK)/Northampton (US)Edgar Elgar Publishing, 2nd ed., 2021, xxviii-634p., pp. 97-112, p. 104; Nicholas Tsagourias, "Electoral Cyber Interference, Self-Determination and the Principle of Non-Intervention in Cyberspace" *in* Dennis Broeders and Bibi van den Berg (eds.), *Governing Cyberspace: Behavior, Power, and Diplomacy*, Lanham/Boulder/New York/London, 2020, Rowman & Littlefield, vii-327p., pp. 45-63.

[38] This contribution is based on the research project *FAKE-ID: Videoanalyse mit Hilfe künstlicher Intelligenz zur Detektion von falschen und manipulierten Identitäten (meaning "AI-based video analysis to detect false and manipulated identities"),* financed by the German Federal Ministry of Education and Research (BMBF) within the framework of the research programme *Künstliche Intelligenz in der zivilen Sicherheitsforschung ("AI in civil security research")* (FKZ: HWR/FÖPS 13N15737, OVGU 13N15736).

[39] See for instance, CBNC.com, "China and Europe are leading the push to regulate A.I. – one of them could set the global playbook" 6 May 2022, at https://www.cnbc.com/2022/05/26/china-and-europe-are-leading-the-push-to-regulate-ai.html.

[40] UNESCO, Recommendation on the ethics of artificial intelligence, November 2021, SHS/BIO/REC-AIETHICS/2021, at https://unesdoc.unesco.org/ark:/48223/pf0000380455.

particular emphasis on a human-centered, ethical, and trustworthy model of AI regulation.[41] Meanwhile, China has recently publicized two legislative drafts that aim to create an ethical framework for regulating AI, as well as one specific proposed bill about deepfakes (called "deep synthesis services" in an unofficial translation).[42] Therefore, there are already emerging and competing regulatory frameworks for AI at national and international level.

In this context, the FAKE-ID project looks at the use of deepfake detection tools by law-enforcement for video-based authentication.[43] Among other research objectives, it aims to react to harmful uses of deepfakes that run counter to EU laws and interests, while ensuring the protection of fundamental rights, democracy, and the rule of law. As we will demonstrate, the overall objective of the FAKE-ID research project is consistent with the nascent European approach on informational digital sovereignty, i.e., "to seek to assert [its] political, economic and social self-determination with regard to digital technology" and to develop its "institutional capacity to reign over developments that affect" societies in the EU.[44]

Various scandals have contributed to making the concept of digital sovereignty more appealing within the European Union, such as the Snowden revelations on US global intelligence practices, the Cambridge Analytica scandal, and the allegations that the 2016 US presidential elections took place under the influence of manipulative data-driven campaigns. We should also not forget the COVID-19 global pandemic. For instance, the presidency of the Council of the European Union stated in its October 2020 conclusions that:

> The COVID-19 pandemic has shown more clearly than ever that Europe must achieve digital sovereignty in order to be able to act with self-determination in the digital sphere and to foster the resilience of the European Union.[45]

This political statement exemplifies the European Union's growing openness towards the necessity to either establish, ensure, or defend its digital sovereignty, including the informational dimensions

---

[41] European Commission, White paper: On Artificial Intelligence – A European approach to excellence and trust, COM (2020) 65 final, 19. February 2020, at https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

[42] chinalawtranslate.com, Provisions on the Administration of Deep Synthesis Internet Information Services (Draft for solicitation of comments), 28 January 2022, at https://www.chinalawtranslate.com/en/deep-synthesis-draft/, Art. 2.

[43] Comp. about deepfake detection for law-enforcement purposes, Europol Innovation Lab, "Facing Reality? Law Enforcement and the Challenge of Deepfakes", 28. April 2022, at https://www.europol.europa.eu/media-press/newsroom/news/europol-report-finds-deepfake-technology-could-become-staple-tool-for-organised-crime; European Parliament, Artificial Intelligence and Law Enforcement: Impact on Fundamental Rights, Studied Requested by the LIBE committee, PE 656.295, July 2020, at https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU(2020)656295_EN.pdf.

[44] Julia Pohle and Daniel Voelsen, "Centrality and Power. The struggle over the techno-political configuration of the Internet and the global digital order", *Policy & Internet*, 2022, Vol. 14, No. 1, pp. 13-27, pp. 20-21.

[45] EU Council, Presidency Conclusions – The Charter of Fundamental Rights in the context of Artificial Intelligence and Digital Change, 11481/20, 21. Oktober 2020, at https://www.consilium.europa.eu/media/46496/st11481-en20.pdf, p. 3. See also, ibid., p. 5 and p. 7.

of the controls and powers that the EU and its member states can exercise over digital forms of information. There are, for instance, growing concerns about the necessity to safeguard the integrity of electoral processes against the rising digital means of influence over political processes. This is also evident in those EU Council conclusions:

> Direct, universal suffrage and free elections by secret ballots are the basis of the democratic process and a core element of our common values. They need to be preserved in the digital era. Cyberattacks and disinformation targeting electoral processes, campaigns and candidates have the potential to polarize public discourse and undermine the secrecy of the ballot, the integrity and fairness of the electoral process and citizens' trust in elected representatives. In this context, we stress the importance of safeguards and active measures to counter disinformation campaigns, the abuse of private data, hybrid threats and cyberattacks.[46]

The very concept of digital sovereignty remains controversial outside of the EU,[47] but it is especially controversial within the EU in terms of what it concretely entails.[48] Despite the lack of a unified European perspective on digital sovereignty, one emerging consensus in the EU equates digital sovereignty internally with the notion of *strategic autonomy*, and externally with the EU's agenda to establish itself as global leader on the basis of its regulatory powers for digital matters and its worldwide influence via the appeal of its standards in related matters: the so-called *Brussels Effect*.[49] Furthermore, there is a trend in the EU towards ensuring some forms of informational privacy and self-determination for peoples and individuals, especially given the increased technology-driven possibilities to exert manipulative influence over societies, in the global process of digitalization.[50]

---

[46] Ibid., p. 13, para. 26.

[47] See for some examples of the how conceptions of sovereignty diverge internationally, Anupam Chander and Haochen Sun, "Sovereignty 2.0", *Vanderbilt Journal of Transnational Law*, 2022? Vol. 55, No. 2, pp. 283-324.

[48] See for instance, Andrej Savin, "Digital Sovereignty and Its Impact on EU Policymaking", *Copenhagen Business School Law Research Paper Series No. 22-02*, 4. April 2022, at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4075106, p. 1; Huw Roberts, Josh Cowls, Frederico Casolari, Jessica Morley, Mariarosaria Taddeo, Luciano Floridi, "Safeguarding European values with digital sovereignty: an analysis of statements and policies", *Internet Policy Review*, 2021, Vol. 10, No. 3, at https://policyreview.info/articles/analysis/safeguarding-european-values-digital-sovereignty-analysis-statements-and-policies, p. 3.

[49] See generally on that concept: Anu Bradford, *The Brussels Effect: how the European Union rules the world*, New York, Oxford University Press, 2020, xix-404p.; Anu Bradford, "The Brussels Effect", *Northwestern University Law Review*, 2012, Vol. 107, No. 1, pp. 1-67. See also, Andrej Savin, "Digital Sovereignty and Its Impact on EU Policymaking", *Copenhagen Business School Law Research Paper Series No. 22-02*, 4. April 2022, at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4075106, p. 4-6, p. 12; Annegret Bendiek and Isabella Stürzer, "Die digitale Souveränität der EU ist umstritten", *SWP-Aktuell 2022/A 30*, April 2022, at https://www.swp-berlin.org/publikation/die-digitale-souveraenitaet-der-eu-ist-umstritten, pp. 5-6; Julia Pohle and Daniel Voelsen, "Centrality and Power. The struggle over the techno-political configuration of the Internet and the global digital order", *Policy & Internet*, 2022, Vol. 14, No. 1, pp. 13-27, pp. 20-21.

[50] See for instance, Anastasia Iliopoulou-Penot, "The construction of a European digital citizenship in the case law of the Court of Justice of the EU", *Common Market Law Review*, 2022, Vol. 59, No. 4, pp. 969-1006.

Indeed, this is one of the core motivations for the European Commission's Proposal for an AI Regulation.[51]

Against this backdrop, this paper focuses on one specific dimension in this overall global context of digital transformation as accelerated by the ongoing artificial intelligence "revolution": the phenomenon of deepfakes and their potential to exert manipulative influence in a way that can affect democratic and security issues. Deepfakes have been defined in a report by the European Parliamentary Research Service titled "Tackling Deepfakes in European Policy" as "manipulated or synthetic audio or visual media that seem authentic, and which feature people that appear to say or do something they have never said or done, produced using artificial intelligence techniques, including machine learning and deep learning."[52]

This paper relates to a research project led by an interdisciplinary research team of IT, law, social and cultural anthropology scholars, working together in a consortium in Germany funded by the German Federal Ministry of Education and Research.[53] This research project, titled "FAKE-ID," aims at researching AI-based detectors for deepfakes in order to contribute to identifying and reacting to deepfakes for security purposes, and also to foster the self-empowerment potential of a deepfake detector for a public willing to assess the authenticity of video or audio content.

One of the main outcomes of the Fake-ID project will be research demonstrators for conducting risk assessments of video-based deepfake threats in authentication applications that will be investigated for identity remote authentication scenarios.[54] Threats for identity proofing are elaborated. A risk and suspicion map as a basis for decision-making is proposed. One further challenge in prioritization and the application of metrics known as factors to compute a Common Vulnerability Scoring System (CVSS) score for a weakness, to include deepfakes in common vulnerability or weakness enumerations[55] are summarized and discussed.

---

[51] See for instance, European Commission, Proposal for a Regulation of the European Parliament and of the Council Laying down harmonized rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts, COM(2021) 206 final, 2021/0106(COD), 21 April. 2021, at https://eur-lex.europa.eu/legal-content/EN-DE/TXT/?from=EN&uri=CELEX%3A52021PC0206, p. 21, (15): "[a]side from the many beneficial uses of artificial intelligence, that technology can also be misused and provide novel and powerful tools for manipulative, exploitative and social control practices."

[52] European Parliamentary Research Service, Tackling deepfakes in European policy, PE 690.039, Juli 2021, at https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf, p. I.

[53] This contribution is based on the research project *FAKE-ID: Videoanalyse mit Hilfe künstlicher Intelligenz zur Detektion von falschen und manipulierten Identitäten (standing for "AI-based video analysis to detect false and manipulated identities")*, financed by the German Federal Ministry of Education and Research (BMBF) within the framework of the research programme *Künstliche Intelligenz in der zivilen Sicherheitsforschung ("AI in civil security research")* (FKZ: HWR/FÖPS 13N15737, OVGU 13N15736).

[54] See for an example of non-deepfakes based threat to remote ID proofing system, Chaos Computer Club, "Chaos Computer Club hackt Video-Ident", 8 August 2022, at https://www.ccc.de/de/updates/2022/chaos-computer-club-hackt-video-ident.

[55] See for further information on the Common Vulnerability Scoring System (CVSS): first.org, "FIRST is the global Forum of Incident Response and Security Teams", at https://www.first.org/cvss/ and the similar cwe.mitre.org, "CWE

2 DEEPFAKE DETECTION AND REMOTE ID PROOFINGS, AS PART OF THE EMERGING UNION APPROACH ON DIGITAL SOVEREIGNTY

Some of the application cases that the FAKE-ID project is looking at concern the use of AI-based tools for detecting deepfakes that can fall under the generic category of remote identity controls or proofing methods (hereafter referred to as "remote ID proofing"). Indeed, remote identity proofing methods "are a way to identify individuals without relying on physical presence."[56] A diverse set of techniques and processes are covered by the formulation that they "can be used in a variety of contexts where trust in the identity of a natural or legal person is essential—such as financial services, e-commerce, travel industry, human resources [and] public administrations".[57]

Remote ID proofing or verification are thus not only performed by public administrations or officials. This is clearly manifest in the growing appeal of such processes of remote ID proofing among private operators such as banks, financial institutions—and in some circumstances, digital service providers.[58] Private actors are also important in this constellation, over and above situations in which they are entrusted to perform remote ID verification, because it is primarily with the help of their datasets that those verification processes are developed and implemented. Indeed, remote ID proofing can be based on several data categories that are collected from various sources and third-party databases—be they private or publicly-owned—which serve as templates or references to verify individuals' identities.[59] This has important implications for data protection and privacy[60] that can have consequences for the compliance of a deepfake detector for law-enforcement purposes with requirements relating to the protection of fundamental rights and the rule of law.[61]

Several methods exist to conduct remote ID proofing. But the approach that is currently most reliable is a combination of several methods, including the use of AI and human intervention to

---

approach ("Common Weakness Enumeration: A Community-Developed List of Software & Hardware Weakness Types")", Page Last Updated on 20 May 2022, at https://cwe.mitre.org/.

[56] ENISA, Remote Identity Proofing: How to spot the Fake from the Real?", 16 July 2021, at https://www.enisa.europa.eu/news/enisa-news/remote-identity-proofing-how-to-spot-the-fake-from-the-real.

[57] Ibid.

[58] Ibid., p. 21.

[59] Ibid., p. 2.

[60] See ibid., pp. 39-40. See also Julia Pohle and Daniel Voelsen, "Centrality and Power. The struggle over the techno-political configuration of the Internet and the global digital order", *Policy & Internet*, 2022, Vol. 14, No. 1, pp. 13-27, p. 22.

[61] Such a deepfakes-detector besides to be subjected to the requirements of Art. 52(3) of the European Commission Proposal for an AI Regulation, will have to respect the requirements applicable to high-risk AI systems under this future Regulation that are set forth under Title III of the Proposal, including data governance practices that must be met by AI providers or also requirements applicable to training-datasets that constitute a core aspect of the development of AI systems. See about that, European Commission, Proposal for a Regulation of the European Parliament and of the Council Laying down harmonized rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts, COM(2021) 206 final, 2021/0106(COD), 21 April. 2021, at https://eur-lex.europa.eu/legal-content/EN-DE/TXT/?from=EN&uri=CELEX%3A52021PC0206, pp. 26-28, (32), (38) and (40) as well as Ibid., Annex III, Sect. 6(d).

operate verifications or final controls. Such mixed approaches are sometimes named "breed methods."[62] The FAKE-ID research project follows a similar mixed approach for detecting deepfakes. The typical outcome of remote ID proofing is the issuance of a proof of authenticity for a person's identity. This can take several forms, such as a confirmation of identity with attribution of an absolute score (YES/NO), a confidence level as a percentage, a likelihood ratio, or the assignment of identification credentials.[63] Similarly, the tools that the Fake-ID research project is investigating to detect deepfakes will generate a score to establish a confidence level that a picture or a video does not constitute a deepfake.

The trend to develop and implement remote ID proofing is rapidly taking off, triggered by the COVID-19 pandemic crisis, which helped establish identity verification without physical presence.[64] Furthermore, this trend is likely to continue in the EU context, given the plan to establish a European digital identity "wallet" common to all EU citizens.[65] The European Commission has publicized a proposal that aims, among other things, to harmonize remote ID proofing of EU identities, both online and offline.[66] In that context, detecting deepfakes will be of increased importance within the EU, given the EU objective to provide "access to highly secure and trustworthy electronic identity solutions" for cross-border activities, so that "that public and private services can rely on trusted and secure digital identity solutions." Another objective of the EU relevant to mention here is empowering and facilitating the use of digital identity solutions by natural and legal persons, including for secure business transactions and access to public services.[67] In its capacity as president of the EU Council in the first half of 2022, France has more recently submitted a second compromise text for the European digital identity wallet, with the primary aim of "prevent[ing] fragmentation of the internal market" and "to define a pan-European legal framework that allows for the cross-border recognition of trust services for the recording of [identity] data in electronic ledgers."[68] One future practical use-

---

[62] ENISA, Remote ID Proofing Analysis of methods to carry out identity proofing remotely, March 2021, at https://www.enisa.europa.eu/publications/enisa-report-remote-id-proofing, p. 25.
[63] Ibid., p. 15.
[64] See for instance, ENISA, Remote ID Proofing Analysis of methods to carry out identity proofing remotely, March 2021, at https://www.enisa.europa.eu/publications/enisa-report-remote-id-proofing, p. 4.
[65] European Commission, European Digital Identity, 28 May 2021, at https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-digital-identity_en#digital-identity-for-all-europeans.
[66] European Commission, Proposal for a Regulation of European Parliament and of the Council amending Regulation (EU) No 910/2014 as regards establishing a framework for a European Digital Identity, 2021/0136(COD), COM(2021) 281 final, 3 June 2021, p. 2.
[67] Ibid., p. 1.
[68] European Commission, Proposal for a Regulation of European Parliament and of the Council amending Regulation (EU) No 910/2014 as regards establishing a framework for a European Digital Identity, 2021/0136(COD), COM(2021) 281 final, 3 June 2021, p. 19, (34) and pp. 23-25. See also for a slightly revised definition of Digital Identity Wallet as amended by the French Presidency of the EU Council, Conseil de l'Union européenne, Proposition de règlement du Parlement européen et du Conseil modifiant le règlement (UE) n°910/2014 en ce qui concerne l'établissement d'un cadre européen relatif à une identité numérique – Deuxième proposition de compromis, 2021/0136(COD), 9200/22, at https://aeur.eu/f/1v7, pp. 19-20, Article 6A.

case of secure access to public services could be to ensure the cyber-security conditions of election infrastructures, including online identity verification for possible forms of e-vote in the future.

The examples show the importance of ensuring secure digital means of identification and identification controls without the need for an individual's physical presence. This will continue to grow in importance and will be taken into account in future regulatory frameworks. In this context, the detection of deepfakes in the overall EU regulatory framework for AI systems will constitute an important aspect of the European digital identity wallet, and will be indirectly relevant for a European conception of digital sovereignty.

3 DEEPFAKES AS POTENTIAL INFORMATIONAL THREATS FOR THE EUROPEAN UNION'S DIGITAL SOVEREIGNTY

The phenomenon of deepfakes potentially affects digital sovereignty as it offers technological means to manipulate digitalized or digital means of identity, including official means of identification.[69] Deepfakes also contribute to contemporary discussions about how informational digital sovereignty can be ensured in light of external "informational threats," with the official aim in the EU to follow a different path than existing illiberal or authoritarian conceptions of informational digital sovereignty. This is a pressing issue due to rapid technological developments that are enabling to generate more and more elaborate deepfakes, but also due to their "democratization" and increasing spread within societies globally.[70]

As mentioned in the introduction, despite the disputed contours of the EU perspective over digital and informational sovereignty, one current minimum consensus within the EU equates sovereignty in the digital context with the objective of ensuring strategic autonomy, mostly against external threats and including informational threats. Against this backdrop, one important constellation for reaching strategic autonomy translates into ensuring cybersecurity, therefore connecting sovereignty and cybersecurity.[71] Indeed, several EU digital policy milestones have emerged in connection with cyber security issues, as shown by the strengthened role attributed to the European Union Agency for Cybersecurity ("ENISA") after the adoption of the EU Cybersecurity

[69] See for some examples in a security-oriented perspective, Europol's European Cybercrime Centre, United Nations Interregional Crime and Justice Research Institute (UNICRI) and Trend Micro, Report on Malicious Uses and Abuses of Artificial Intelligence (AI), 19. November 2020, at https://eucrim.eu/news/report-on-malicious-uses-and-abuses-of-artificial-intelligence/, pp. 54-65.

[70] See for instance in the scientific context, Chemistryworld.com, "AI-generated images could make it almost impossible to detect fake papers", 24 May 2022, at https://www.chemistryworld.com/news/ai-generated-images-could-make-it-almost-impossible-to-detect-fake-papers/4015708.article.

[71] Andrej Savin, "Digital Sovereignty and Its Impact on EU Policymaking", *Copenhagen Business School Law Research Paper Series No. 22-02*, 4. April 2022, at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4075106, p. 4.

Regulation in 2019.[72] Since the adoption of the 2019 EU Cybersecurity Act, ENISA has been entrusted among others with the task of "contribut[ing] to the development and implementation of Union policy and law," by "supporting […] the development and implementation of Union policy in the field of electronic identity and trust services."[73]

In this context, one important question is whether deepfakes can constitute a cybersecurity threat to an emerging EU digital sovereignty conception. This question requires a nuanced reflection, for three reasons. First, various application cases already exist for deepfakes, but it is important to note that deepfakes can also be used for artistic,[74] political,[75] educational,[76] entertainment,[77] or medical purposes.[78] Second, there are two constellations wherein deepfakes can clearly deepen "cybersecurity threats," namely disinformation and identity manipulations/thefts. Third, both constellations can interrelate in the particular context of elections, where digital identity plays a crucial role: Either (i) they can be used in formal contexts, to verify the authenticity of nationals allowed to vote for an election, or (ii) in an informal way, to prevent online contents from being spread online via accounts using fake identities to manipulate electoral processes in the increasingly digitalized dimensions of public debates.[79] Admittedly, the second constellation poses delicate issues given the anonymity that users of online platforms can often enjoy.

Furthermore, protection against identity manipulation and thefts refers to the general concept of sovereignty, insofar as it deeply relates to the concept of nationality and to the emerging legal concept of digital citizenship in EU law. Under general international law, one of the traditional core prerogatives of states is to attribute nationality to individuals and verify it. However, there is "no

---

[72] Annegret Bendiek and Isabella Stürzer, "Die digitale Souveränität der EU ist umstritten", *SWP-Aktuell 2022/A 30*, April 2022, at https://www.swp-berlin.org/publikation/die-digitale-souveraenitaet-der-eu-ist-umstritten, pp. 3-4.

[73] European Parliament and EU Council, Regulation (EU) 2019/881 on ENISA (the European Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act), PE/86/2018/REV/1, Art. 5.

[74] See for instance, wired.co.uk, "These historical artefacts are totally faked", 24 October 2021, at https://www.wired.co.uk/article/fake-artefacts-ai.

[75] See for instance, Umur A. Ciftci, Gokturk Yuksek, Ilke Demir, "My Face My Choice: Privacy Enhancing Deepfakes for Social Media Anonymisation", 2 November 2023, at https://arxiv.org/pdf/2211.01361v1.pdf.

[76] See for instance, Wired.com, "Deepfakes Are Becoming the Hot New Corporate Training Tool", 7 July 2020, at https://www.wired.com/story/covid-drives-real-businesses-deepfake-technology/.

[77] See for instance, Ft.com, "Deepfakes: Hollywood's quest to create the perfect digital human", 10 October 2019, at https://www.ft.com/content/9df280dc-e9dd-11e9-a240-3b065ef5fc55.

[78] European Parliamentary Research Service, Tackling deepfakes in European policy, PE 690.039, Juli 2021, at https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf.

[79] See the eight scenarios developed for illustrating ethical harms that could be generated by the use of deepfakes in the context of elections, Nicholas Diakopoulos and Deborah Johnson, "Anticipating and addressing the ethical implications of deepfakes in the context of elections", *New Media & Society*, 2021, Vol. 23, No. 7, pp. 2072-2098. See also Tom Dobber, Nadia Metoui, Damian Trilling, Nathalie Helberger, Claes de Vreese, "Do (Microtargeted) Deepfakes Have Real Effects on Political Attitudes?", *The International Journal of Press/Politics*, 2021, Vol. 26, No. 1, pp. 69-91. See about allegations regarding the use of false identity online in the context of the 2016 U.S. presidential elections, Michael N. Schmitt, ""Virtual" Disenfranchisement: Cyber Election Meddling in the Grey Zones of International Law", *Chicago Journal of International Law*, 2018, Vol. 19, No. 1, pp. 30-67, p. 36.

coherent, accepted definition of nationality in international law and only conflicting descriptions under the different municipal laws of states." Furthermore, "the rights and duties attendant upon nationality vary from state to state."[80] That said, nationality still constitutes a core link between states and "their" peoples and therefore directly involves the concept of sovereignty.[81] This principle was confirmed by the Permanent Court of International Justice in its 1923 *Nationality Decrees in Tunis and Morocco* case, in which it was decided that "it is for each state to determine under its own law who are its nationals," before adding that this "law shall be recognized by other states in so far as [applicable international law] with regard to nationality."[82] In its 1955 *Nottebohm* case, the International Court of Justice established that under international law, nationality with respect to the State granting it, is "a legal bond having as its basis a social fact of attachment, a genuine connection of existence, interests and sentiments, together with the existence of reciprocal rights and duties."[83] If international law—and in particular international human rights law—has restricted the freedom of States to attribute and control the nationality of persons under its jurisdiction since the 1950s, this still remains the position of principle under currently applicable international law:

> Even if the freedom of States to regulate their nationality is much more restricted today, considering the development of international law since 1923, that [Permanent International Court of Justice's] statement is essentially still valid: each State is in principle still entitled to determine under its own law who are its nationals (cf. Art. 3 (1) European Convention on Nationality). International law limits that discretion, but it neither contains nor prescribes certain criteria for acquisition and loss of nationality.[84]

Neither processes of digitalizing traditional forms of identification documents and establishing digital forms of identification documents and controls fundamentally disturb this core relation between a

[80] Malcom N. Shaw, *International Law*, 2014, Cambridge, CUP, 7th Edition, 1063p., p. 479.
[81] Matthias Leese, "Fixing State Vision: Interoperability, Biometrics, and Identity Management in the EU", *Geopolitics*, 2022, Vol. 27, No. 1, pp. 113-133, p. 114.
[82] Permanent Court of International Justice, *Nationality Decrees in Tunis and Morocco Case*, Series B, No. 4, 1923; 2 AD, p. 24.
[83] International Court of Justice, Nottebohm Case (second phase) (Lichtenstein v. Guatemala), Judgment of April 16th, 1955, I.C.J. Reports, p. 4, p. 23.
[84] Oliver Dörr, "Nationality", *Max Planck Encyclopedia of International Law*, August 2019, para. 4.

state's sovereignty and personal jurisdiction[85] as exercised over nationals from that state,[86] subject to restrictions imposed by international law.

Identity manipulation is increasingly perceived in the EU as a potential threat to its sovereignty, because it could endanger EU laws, interests, and values if it materialized at a general level.[87] Such a claim might appear exaggerated in the contemporary context, but as digital forms of identification processes and related verification mechanisms gradually grow, so too will such threats, especially deepfake-based threats. Even if such major deepfake-based threats have not yet made their presence felt,[88] there is a serious possibility, evidenced in the many regulatory efforts to counter them, that security threats may rapidly increase, since deepfakes are increasingly easier to create and use.[89] The logic increasingly at play here that links digital sovereignty and identity control in the process of digitalization[90] can indeed be broadly compared to the phenomenon of smart borders,[91] despite the fact that remote ID proofing is meant to apply independently of any proper physical or digitalized borders and does not focus on the control and identification of foreigners. Several legal instruments have already been adopted with that aim at the EU level, that is, in order to regulate how digital forms of identity can be established and controlled in the EU without the individuals in question being physically present. That said, these developments can also be critically assessed, as they for instance arguably participate to frame the figure of the "foreigner" as an indirect threat to the EU sovereignty, especially given wide-spread securitization discourses in EU policies and legislations.[92]

---

[85] In this sense, see for instance Onuma Yasuaki, *International Law in a Transcivilizational World*, Cambridge, Cambridge University Press, 2017, 666p., p. 333: "This bond constitutes the basis for a state's jurisdiction over its members (so-called personal sovereignty or jurisdiction). A state can apply its law over tis nationals even when they are outside of its territory, including within the territory of a foreign state […]. Furthermore, while nationality is fundamentally a concept concerning natural persons, it plays an intermediary function in the application of jurisdiction over juridical persons, whose activities transcend national borders."

[86] Comp. with Ewa Michalkiewicz-Kadziela, Ewa Milczarek, "Legal boundaries of digital identity creation", *Internet Policy Review*, 2022, Vol. 11, No. 1, at https://policyreview.info/articles/analysis/legal-boundaries-digital-identity-creation.

[87] See for instance the concept of Foreign Information Manipulation and Interference (FIMI) used by some European agencies, ENISA, Foreign information manipulation and interference (FIMI) and cybersecurity – Threat Landscape, December 2022, at https://www.enisa.europa.eu/news/cybersecurity-foreign-interference-in-the-eu-information-ecosystem.

[88] See for instance, Trend Micro, "How Underground Groups Use Stolen Identities and Deepfakes", 27 September 2022, at https://www.trendmicro.com/en_us/research/22/i/how-underground-groups-use-stolen-identities-and-deepfakes.html.

[89] See for instance, U.S. Congressional Research Service, "Deep Fakes and National Security", Updated on 3 June 2022, at https://crsreports.congress.gov/product/pdf/IF/IF11333.

[90] See on that interrelation, Matthias Leese, "Fixing State Vision: Interoperability, Biometrics, and Identity Management in the EU", *Geopolitics*, 2022, Vol. 27, No. 1, pp. 113-133, esp. pp. 116-120.

[91] See on that phenomenon, Ayelet Shachar, *The Shifting Border: Legal Cartographies of Migration and Mobility*, Manchester University Press, 2020, 328 p.; Jonas Püschmann, Book Review: The Shifting Border: Legal Cartographies of Migration and Mobility, 18 March 2022, at https://blogs.law.ox.ac.uk/research-subject-groups/centre-criminology/centreborder-criminologies/blog/2022/03/book-review.

[92] See for instance, euronews.com, "Joseph Borell apologises for controversial 'garden vs jungle' metaphor but defends speech", 20 November 2022, at https://www.euronews.com/my-europe/2022/10/19/josep-borrell-apologises-for-controversial-garden-vs-jungle-metaphor-but-stands-his-ground.

Remote ID verification increasingly relates to the emerging EU approach for digital sovereignty, as understood in its minimalistic conception that relate to strategic autonomy and cyber security concerning so-called informational threats. Art. 24(1) of Regulation (EU) 910/2014 on electronic identification and trust services ("eIDAS")[93] is a good example of how remote ID proofing *already relates* to the exercise of states' sovereignty through digital means, by relativizing the traditional importance of physical powers exercised territorially by sovereign states in the international society. This provision mandates that:

> [w]hen issuing a qualified certificate for trust service, a qualified certificate for a trust service provider shall verify, by appropriate means and in accordance with national law, the identity and, if applicable, any specific attributes of the natural or legal person to whom the qualified certificate is issued.[94]

The link between remote ID proofing and sovereignty is strengthened in some use cases for which EU law requires a verification of the identity of persons in online transactions for the purposes of countering money laundering or terrorism financing, as foreseen by the Anti-Money Laundering/Counter Financing Terrorism (AMT/CFT) directives. The fifth AML/CFT Directive was adopted to also strengthen the possibilities for the EU to monitor financial transactions, including regarding the identity of persons involved in those transactions, especially with respect to third countries that are regarded as a source of risk due to an insufficient level of control over money laundering and terrorism financing.[95] Art. 9 of the fifth AMD/CFT Directive seeks, for instance, to protect the integrity of the European financial system, which also relates to an emerging minimal understanding of European digital sovereignty.[96] More generally, the exploitation of cyber security vulnerabilities via the manipulation of identities[97] can lead to the emergence of threats that are not

---

[93] European Parliament and EU Council, Regulation (EU) 910/2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC, *Official Journal of the European Union L 257/73*, 28 August 2014.

[94] European Parliament and EU Council, Regulation (EU) 910/2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC, *Official Journal of the European Union L 257/73*, 28 August 2014, p. 26, Art. 24(1).

[95] European Commission, "Strengthened EU rules to prevent money laundering and terrorism financing" (Fact sheet), 9 July 2018, at https://ec.europa.eu/info/files/factsheet-main-changes-5th-anti-money-laundering-directive_en.

[96] European Parliament and EU Council, Directive (EU) 2015/849 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing, amending Regulation (EU) No 648/2012 of the European Parliament and of the Council, repealing Directive 2005/60/EC, 20 May 2015, p. 18, Art. 9(1). Comp. with Andrej Savin, "Digital Sovereignty and Its Impact on EU Policymaking", *Copenhagen Business School Law Research Paper Series No. 22-02*, 4. April 2022, at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4075106, p. 2.

[97] See for some real examples of cyber security threats based on manipulated identities, Trend Micro, "How Underground Groups Use Stolen Identities and Deepfakes", 27 September 2022, at https://www.trendmicro.com/en_us/research/22/i/how-underground-groups-use-stolen-identities-and-deepfakes.html.

only damaging for parties involved in an online transaction or communication but more broadly for a country, if a sufficient gravity threshold is reached.[98]

The fact that consensus over digital sovereignty within the EU is mostly found in relation to the notion of strategic autonomy and cybersecurity threats is illustrated by several recent EU Council conclusions that all put the emphasis on the importance of cybersecurity and informational self-determination for the EU approach on digital sovereignty.[99] Developments at the EU level clearly indicate a willingness to move forward with a common cybersecurity strategy serving the whole European Union's digital and informational sovereignty. Indeed, they underscore the fact that securing digital means of identity as well as cybersecurity processes aiming specifically at the protection of the integrity of decision-making processes are increasingly influential in the emergence of a minimal European understanding of digital sovereignty.

4 CONCLUSION

In this context, deepfake detection exerts a narrow but nonetheless important role. Indeed, deepfakes are increasingly perceived as being able to threaten decision-making processes in the global context of digitalization, while posing threats to the security of persons and societies within the European Union. For this reason, deepfake detection and its use for remote ID verification integrates the emerging Union approach over digital informational sovereignty, which is for the moment mostly focused on ensuring security and strategic autonomy, while protecting fundamental rights, democracy and the rule of law.

5 ACKNOWLEDGMENTS

The work in this paper is funded in part by the German Federal Ministry of Education and Research (BMBF) under FAKE-ID project: grant numbers OVGU FKZ: 13N15736 and HWR/FPÖS FKZ: 13N15737. We would like to thank all project partners for the fruitful discussion and exchange in the project.

[98] ENISA, Remote ID Proofing Analysis of methods to carry out identity proofing remotely, March 2021, at https://www.enisa.europa.eu/publications/enisa-report-remote-id-proofing, pp. 45-46.

[99] EU Council, Council conclusions on Foreign Informational Manipulation and Interference (FIMI), 11429/22, 18 July 2022, at https://data.consilium.europa.eu/doc/document/ST-11429-2022-INIT/en/pdf; EU Council, Council conclusions on a Framework for a coordinated EU response to hybrid campaigns, 10016/22, 21 June 2022, at https://data.consilium.europa.eu/doc/document/ST-10016-2022-INIT/en/pdf; EU Council, Council conclusions on the Special Report of the European Court of Auditors No 05/2022 entitled 'Cybersecurity of the EU Institutions, bodies and agencies: Level of preparedness overall not commensurate with the threats, 10504/22, 21 June 2022, at https://data.consilium.europa.eu/doc/document/ST-10016-2022-INIT/en/pdf.

We thank the participants in our panel during the 2022 Weizenbaum Institute Conference and we would like furthermore to thank Prof. Harmut Aden (HWR) for his comments and suggestions, as well as Mario Petoshati (HWR) for his thorough review and editing.

6 REFERENCES

1. Appel M., Prietzel F. (2022). The detection of political deepfakes. Journal of Computer-Mediated Communication, 27(4), at https://academic.oup.com/jcmc/article/27/4/zmac008/6650406 (visited on 7 February 2023).
2. Bendiek A., Stürzer I. (2022). Die digitale Souveränität der EU ist umstritten. SWP-Aktuell 2022/A 30. At https://www.swp-berlin.org/publikation/die-digitale-souveraenitaet-der-eu-ist-umstritten (visited on 29 August 2022).
3. Bodi M. (2021), The First Amendment Implications of Regulating Political Deepfakes. Rutgers Computer and Technology Law Journal, 47(1), 143-172.
4. Blitz M. J. (2020), Deepfakes and Other Non-Testimonial Falsehoods: When is Belief Manipulation (Not) First Amendment Speech? Yale Journal of Law & Technology, 23(3), 160-300.
5. Bradford A. (2020). The Brussels Effect: how the European Union rules the world. New York, Oxford University Press, xix-404p.
6. Bradford A. (2012). The Brussels Effect. Northwestern University Law Review, 107 (1), 1-67.
7. CBNC.com, "China and Europe are leading the push to regulate A.I. – one of them could set the global playbook"6 May 2022, at https://www.cnbc.com/2022/05/26/china-and-europe-are-leading-the-push-to-regulate-ai.html.
8. Chander A., Sun H. (2022). Sovereignty 2.0. Vanderbilt Journal of Transnational Law, 55 (2), 283-324.
9. Chesney B., Citron D. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. California Law Review, 107(6), 1753-1820.
10. Chaos Computer Club (2022). Chaos Computer Club hackt Video-Ident, 8 At https://www.ccc.de/de/updates/2022/chaos-computer-club-hackt-video-ident (page visited on 28 August 2022).
11. Chemistryworld.com (2022). AI-generated images could make it almost impossible to detect fake papers. At https://www.chemistryworld.com/news/ai-generated-images-could-make-it-almost-impossible-to-detect-fake-papers/4015708.article (page visited on 28 August 2022).
12. chinalawtranslate.com, Provisions on the Administration of Deep Synthesis Internet Information Services (Draft for solicitation of comments), 28 January 2022, at https://www.chinalawtranslate.com/en/deep-synthesis-draft/.
13. Ciftci U. A., Yuksek G., Demir I (2023). My Face My Choice: Privacy Enhancing Deepfakes for Social Media Anonymisation, at https://arxiv.org/pdf/2211.01361v1.pdf.
14. Conseil de l'Union européenne (2021), Proposition de règlement du Parlement européen et du Conseil modifiant le règlement (UE) n°910/2014 en ce qui concerne l'établissement d'un cadre européen relatif à une identité numérique – Deuxième proposition de compromis. 2021/0136(COD), 9200/22.
15. cwe.mitre.org (2022). CWE approach ("Common Weakness Enumeration: A Community-Developed List of Software & Hardware Weakness Types"). At https://cwe.mitre.org/ (page visited on 28 August 2022).
16. Diakopoulos N., Johnson D. (2021). Anticipating and addressing the ethical implications of deepfakes in the context of elections. New Media & Society, 23(7), 2072-2098.
17. Dobber T., Metoui N., Trilling D., Helberger N., de Vreese C. (2021). Do (Microtargeted) Deepfakes Have Real Effects on Political Attitudes?. The International Journal of Press/Politics, 26(1), 69-91.

18. Dörr (2019). Nationality. Max Planck Encyclopedia of International Law.
19. ENISA (2021). Remote Identity Proofing: How to spot the Fake from the Real?. At https://www.enisa.europa.eu/news/enisa-news/remote-identity-proofing-how-to-spot-the-fake-from-the-real (page visited on 28 August 2022).
20. ENISA (2022). Foreign information manipulation and interference (FIMI) and cybersecurity – Threat Landscape. At https://www.enisa.europa.eu/news/cybersecurity-foreign-interference-in-the-eu-information-ecosystem.
21. euronews.com (2022), Joseph Borell apologises for controversial 'garden vs jungle' metaphor but defends speech. At https://www.euronews.com/my-europe/2022/10/19/josep-borrell-apologises-for-controversial-garden-vs-jungle-metaphor-but-stands-his-ground.
22. European Commission (2021)(a). Proposal for a Regulation of European Parliament and of the Council amending Regulation (EU) No 910/2014 as regards establishing a framework for a European Digital Identity. 2021/0136(COD), COM(2021) 281 final.
23. European Commission (2021)(b). European Digital Identity. At https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-digital-identity_en#digital-identity-for-all-europeans (page visited on 28 August 2022).
24. European Commission (2021)(c). Proposal for a Regulation of the European Parliament and of the Council Laying down harmonized rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. COM(2021) 206 final, 2021/0106(COD).
25. European Commission (2018). Strengthened EU rules to prevent money laundering and terrorism financing" (Fact sheet). At https://ec.europa.eu/info/files/factsheet-main-changes-5th-anti-money-laundering-directive_en (page visited on 28 August 2022).
26. European Commission (2020). White paper: On Artificial Intelligence – A European approach to excellence and trust. COM (2020) 65 final. At https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf (visited on 28 August 2022).
27. EU Council (2020) Presidency Conclusions – The Charter of Fundamental Rights in the context of Artificial Intelligence and Digital Change. 11481/20. At https://www.consilium.europa.eu/media/46496/st11481-en20.pdf.
28. European Parliament (2020). Artificial Intelligence and Law Enforcement: Impact on Fundamental Rights, Studied Requested by the LIBE committee, PE 656.295. At https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU(2020)656295_EN.pdf.
29. European Parliament and EU Council (2019). Regulation (EU) 2019/881 on ENISA (the European Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act). PE/86/2018/REV/1.
30. European Parliament and EU Council (2015). Directive (EU) 2015/849 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing, amending Regulation (EU) No 648/2012 of the European Parliament and of the Council, repealing Directive 2005/60/EC.
31. European Parliament and EU Council (2014). Regulation (EU) 910/2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive *1999/93/EC. Official Journal of the European Union L 257/73*.

32. European Parliamentary Research Service (2021). Tackling deepfakes in European policy. PE 690.039. At https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf (page visited on 28 August 2022).
33. Europol's European Cybercrime Centre, United Nations Interregional Crime and Justice Research Institute (UNICRI) and Trend Micro (2020). Report on Malicious Uses and Abuses of Artificial Intelligence (AI). At https://eucrim.eu/news/report-on-malicious-uses-and-abuses-of-artificial-intelligence/.
34. Europol Innovation Lab (2022. Facing Reality? Law Enforcement and the Challenge of Deepfakes. At https://www.europol.europa.eu/media-press/newsroom/news/europol-report-finds-deepfake-technology-could-become-staple-tool-for-organised-crime.
35. first.org (2022). FIRST is the global Forum of Incident Response and Security Teams. At https://www.first.org/cvss/ (page visited on 28 August 2022).
36. Ft.co (2019). Deepfakes: Hollywood's quest to create the perfect digital human. At at https://www.ft.com/content/9df280dc-e9dd-11e9-a240-3b065ef5fc55 (page visited on 28 August 2022).
37. Iliopoulou-Penot A. (2022). The construction of a European digital citizenship in the case law of the Court of Justice of the EU. Common Market Law Review, 59 (4), 969-1006.
38. International Court of Justice (1955). Nottebohm Case (second phase) (Lichtenstein v. Guatemala). Judgment of April 16th, 1955, I.C.J. Reports, p. 4.
39. Kivovaty, I. (2021). The international law of cyber intervention *in* Tsagourias N. and Buchan R. (eds.), *Research Handbook on International Law and Cyberspace* (Cheltenham (UK)/Northampton (US); Edgar Elgar Publishing: 2021), 97-112.
40. Leese (M.) (2022). Fixing State Vision: Interoperability, Biometrics, and Identity Management in the EU. Geopolitics, 27(1), 113-133.
41. Meta (2020). Enforcing Against Manipulated Media. At https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/ (page visited on 28 August 2022).
42. Michalkiewicz-Kadziela E., Milczarek E. (2022). Legal boundaries of digital identity creation. Internet Policy Review, 11 (1).
43. reseach.google.com, "Colaboratory: Frequently Asked Questions", at https://research.google.com/colaboratory/faq.html (page visited on 28 August 2022).
44. Roberts H., Cowls J., Casolari F., Morley J., Taddeo M., Floridi F. (2021). Safeguarding European values with digital sovereignty: an analysis of statements and policies. Internet Policy Review, 10 (3), at https://policyreview.info/articles/analysis/safeguarding-european-values-digital-sovereignty-analysis-statements-and-policies (visited on 28 August 2022).
45. reuters.com, "Exclusive: Google, Facebook, Twitter to tackle deepfakes or risk EU fines", 14 June 2022, at https://www.reuters.com/technology/google-facebook-twitter-will-have-tackle-deepfakes-or-risk-eu-fines-sources-2022-06-13/.
46. Permanent Court of International Justice (1923). Nationality Decrees in Tunis and Morocco Case, Series B, No. 4, 1923; 2 AD.

47. Pohle J., Voelsen D. (2022). Centrality and Power. The struggle over the techno-political configuration of the Internet and the global digital order. Policy & Internet, 14 (1), 13-27.

48. Savin A. (2022). Digital Sovereignty and Its Impact on EU Policymaking. Copenhagen Business School Law Research Paper Series No. 22-02, , at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4075106 (page visited on 28 August 2022).

49. Schmitt M. N. (2018). "Virtual" Disenfranchisement: Cyber Election Meddling in the Grey Zones of International Law. Chicago Journal of International Law, 19 (1), 30-67.

50. Shaw M. N. (2014) International Law. 2014, 7th Edition, Cambridge, Cambridge University Press, 1063p..

51. TechRadar.com, "Google is cracking down hard on deepfakes", 31 May 2022, at https://www.techradar.com/news/google-is-cracking-down-hard-on-deepfakes.

52. Trend Micro (2022). How Underground Groups Use Stolen Identities and Deepfakes. At https://www.trendmicro.com/en_us/research/22/i/how-underground-groups-use-stolen-identities-and-deepfakes.html.

53. Tsagourias N. (2020). "Electoral Cyber Interference, Self-Determination and the Principle of Non-Intervention in Cyberspace" in Broeders D. and van den Berg B. (eds.), Governing Cyberspace: Behavior, Power, and Diplomacy (Lanham/Boulder/New York/London; Rowman & Littlefield: 2020), 45-63.

54. UNESCO, Recommendation on the ethics of artificial intelligence, November 2021, SHS/BIO/REC-AIETHICS/2021, at https://unesdoc.unesco.org/ark:/48223/pf0000380455.

55. U.S. Congressional Research Service (2022). Deep Fakes and National Security. At https://crsreports.congress.gov/product/pdf/IF/IF11333 (page visited on 28 August 2022).

56. wired.co.uk (2021). These historical artefacts are totally faked. At https://www.wired.co.uk/article/fake-artefacts-ai (page visited on 28 August 2022).

57. Wired.com (2020).Deepfakes Are Becoming the Hot New Corporate Training Tool At https://www.wired.com/story/covid-drives-real-businesses-deepfake-technology/ (28 August 2022).

58. Yasuaki O. (2017). International Law in a Transcivilizational World, Cambridge, Cambridge University Press.

Proceedings of the Weizenbaum Conference 2022:
Practicing Sovereignty. Interventions for Open Digital Futures

DIGITAL COMMONS AS A MODEL FOR DIGITAL SOVEREIGNTY

THE CASE OF CULTURAL HERITAGE

Lehmann, Jörg
Berlin State Library
Berlin, Germany
joerg.lehmann@sbb.spk-berlin.de

KEYWORDS

commons; cultural heritage; digitization; big data; commodification; sustainability

DOI: 10.34669/wi.cp/4.15

ABSTRACT

This contribution looks at cultural heritage institutions and their digital assets from a commons perspective. Since the beginning of digitization in the late 1990s and with the change of the medium from the analogue to the digital, the role and mission of cultural heritage institutions has changed. Challenges for managing their assets in the sense of a commons arise, on the one hand, due to the current legislation on copyright and intellectual property rights, and, on the other, because of the availability of digital cultural heritage as Big Data, which opens up possibilities for economic exploitation of these assets by private companies. Should digital assets be available open access, or should access and use be regulated? This short paper discusses the possibilities for this model of sovereign data governance within the legal regimes of intellectual property rights and the public domain.

1 INTRODUCTION

A (digital) commons is a shared good or resource that is managed by a community for the benefit of its members, or, in a broader sense, is accessible for society or even for the global population. Digital cultural heritage can be understood as a commons, since digitization has mostly been funded by the public sector and because it is available open access via the internet. Such an understanding of digital cultural heritage as a commons is fostered by Article 27 of the Universal Declaration of Human Rights, which says that everyone has the right to freely participate in the cultural life of the community. The whole of cultural heritage institutions can therefore be conceptualized as the community governing and managing digital resources for the common welfare. Just like commons institutions, they neither pertain to the market nor to the state. The emphasis here is rather on sovereignty and self-organization, but the state—or in this case the European Union—provides the regulatory framework. The concept of the commons has been adapted in the 21st century to apply to the digital age, and the characteristics of a digital cultural commons have been carved out (Haux, 2021). A framework for the analysis and systematic comparison of commons institutions has been developed by Frischmann, Madison, & Strandburg (2014). Since the 1990s, millions of items of cultural heritage have been digitized. However, in many cases, cultural heritage institutions are not allowed to provide access to everything that is available in digital form; intellectual property rights prohibit this. Therefore, in a conventional understanding, digital cultural heritage refers to works which are in the public domain. Commercial use and the free re-use of digital assets which are not in the public domain are excluded by default. These restrictions therefore restrain the use of digital assets for the purpose of creating culture anew. Furthermore, the scale of digitization has turned cultural heritage into a commodity; such vast digital assets have a value as Big Data, since they can be used for machine learning applications, for machine translation, or the establishment of large language models. Now the question arises of who would benefit from this value: whether it serves the interests of private companies interested in optimizing their services and maximizing their profits, or whether it serves the common good. This issue has several twists: Because digital assets can be copied endlessly without the risk of the resource becoming exhausted, it is impossible to over-use the digital resource. However, there could be a potential loss of communal benefits due to actions motivated by self-interest (Yakowitz Bambauer, 2011): Private companies, for example, are not members of the commons, and the profits they might create out of the assets digitized mostly with taxpayers' money might not flow back into the commons. Such commercial use may preclude cultural heritage institutions from tapping into the potential of value creation and impair their digital sovereignty in managing the access to the digital assets as well as with regard to the maintenance of the commons.

The questions posed here have been discussed in a series of interviews with a range of cultural heritage practitioners and with law scholars, using the methodology developed by Frischmann et al. This short paper presents some of the key insights of a research project conducted at the Center for Advanced Internet Studies (CAIS, Bochum) during the winter term 2021/2022, the results of which were published, alongside the transcribed interviews, by Lehmann (2022).

2 DIGITAL SOVEREIGNTY WITHIN THE CURRENT LEGAL FRAMEWORK

Cultural heritage institutions fulfil an important task by selecting objects and collections from the vast pool of cultural products to preserve them and to provide access. These institutions therefore play a crucial role in defining what cultural heritage is and what of the totality of cultural products is going to be preserved. The value of cultural heritage is created by the expert knowledge and the procedures centered in the institutions that are responsible for identifying the cultural value of cultural goods—be it historical, artistic, scientific, architectural, archaeological, or otherwise—or that evaluate the meaning of a particular piece or collection for a specific community. Irrespective of whether physical objects or intangible cultural heritage are in the focus, the selection performed by cultural heritage institutions initiates a process of musealization and decontextualization (Lenski, 2013). The perspective shifts from the function of a good within its specific cultural context to the preservation of a tradition that is regarded as a form of cultural expression. Cultural heritage institutions perform the task of selection based on their expertise and by deploying the procedures their personnel—such as curators, archivists, librarians, conservators, or researchers—have learned in their specialist education. Galleries, libraries, archives, and museums (also called GLAM institutions) can often look back on a long pedigree and are endowed with high reputation and trust, which provides a certain quality assurance in the selection of the objects. By placing cultural products in these institutions, an ennoblement as cultural heritage is taking place. This function of valorization performed by cultural heritage institutions can be contrasted with the acquisitiveness of the big tech companies, which tend to collect each and every digital asset without any further differentiation and store them as Big Data in their data warehouses, and which do not have the means and procedures to address challenges central to the selection process and the forming of collections (Jo & Gebru, 2020). To a certain extent, interaction with the free market is evident, especially with the art market, which ranges from auction houses to the antiques trade; examples here include paintings and miniatures from art history, which are bought both by museums as well as by private collectors. The value of digital cultural heritage is indicated by its availability as Big Data. The vastness of the resources that are now available offers opportunities for economic exploitation for both private companies and

cultural heritage institutions, for example, by constructing large language models used for the improvement of their services, the attraction of more users and for increasing their revenues. The establishment of such large language models does not only raise the question of the ecological burden of computation and therefore of sustainability, but also of the consequences of the biases by which they are marked (Bender, Gebru, McMillan-Major & Shmitchell, 2021). In contrast to big tech companies, cultural heritage institutions have an excellent knowledge of the sources and domains from which the content they work with comes, they have metadata at hand which enable its careful curation, and they may therefore be able to provide high-quality products that consume less energy and serve their societies better than the models established by private companies would do (Lehmann, 2022).

The current legal framework in which cultural heritage institutions operate is mainly marked by the two rights regimes of public domain and intellectual property rights, both of which apply to the works under consideration here. The intellectual property rights regime has to be understood as a complement to the public domain part of cultural heritage, or, as James Boyle has put it, the public domain and the idea of the commons form the outside of intellectual property (Boyle, 2008). However, there are several transient zones between the two rights regimes, and consequently, there are several legal insecurities that arise out of the question how to deal with such material (like, for example, orphaned works, grey literature, leaflets and broadsheets etc.). But generally speaking, and in a conventional understanding, digital cultural heritage refers to works that are in the public domain.

In current public law, the purpose of cultural heritage is described as protecting and valorizing cultural traditions, but the aim of advancing cultural development is also cited (Lenski, 2013). This understanding of the function of cultural heritage was formed by the pre-digital age, where new cultural works were created through reception, be it through reading a book in a library or visiting a museum. The availability of cultural heritage in digital form, however, has changed the relationship between cultural heritage institutions and their users in multiple ways: Users have become accustomed to working with digital material that is in the public domain, but they are also interested in getting access to digital assets that have been produced in the past 70 years and may therefore be protected by intellectual property rights. This rights regime conflicts with the cultural practices established in communities in the creative sector who work with digital material and with their expectations regarding the open accessibility of such digital assets. Moreover, digital reproductions facilitate new modes of reuse that result out of qualities specific to the digital. Good examples here include the remixing of music, the animation of images or their conversion from 2D into 3D, and the creation of multi-modal books in electronic formats. Mission creep can be noted with regard to cultural heritage institutions: In comparison to the pre-digital age, their emphasis is no longer only

on preserving cultural heritage and providing access to it, but also on enabling the re-use of the cultural products which are available in digital format—with the aim of creating culture anew. A central challenge for cultural heritage institutions therefore consists in making available digital assets that are protected by intellectual property rights, with the purpose of stimulating the creation of culture. In so doing, cultural institutions would be managing their digital assets in a sovereign way.

Cultural heritage institutions have a range of possibilities and tools in this respect: They can negotiate with legators and the rights holders of the legacies to enable the re-use of material produced in the 20th century and of born-digital contents; this re-use may only be granted to registered users under certain conditions and in restricted spaces and not, as is current practice, to every possible user worldwide. Furthermore, cultural heritage institutions can engage with communities already working with digital assets; they can recognize their cultural practices as collective customs and traditions. An example of this has been given by the German UNESCO commission which has granted the status of intangible cultural heritage to the Demo scene, a subcultural computer art movement marked by comparably long traditions and customs (German Commission for UNESCO, 2021). Finally, cultural heritage institutions can invite such communities to create, curate, and pool their resources as digital cultural heritage and ask these communities to provide access to such digital resources, be it in the sense of open access for everyone or by enabling re-use of this material under certain conditions or with restricted access. In a certain way, the establishment of such relationships between cultural heritage institutions and the users of the digital assets they provide resembles a classical and historical conception of commons as closed spaces that contain resources to which only an elite has access. However, a trade-off has to be noted: While cultural heritage institutions can engage in sovereign management of their digital assets, they have to restrict access to them by excluding nonregistered users.

The second challenge for digital cultural heritage—the possible exploitation of digital cultural heritage as big data by private companies—can be addressed by using the possibilities given in the current legal framework. The European Directive on Copyright in the Digital Single Market (European Commission, 2019) has introduced two mandatory exceptions for cultural heritage institutions. The first exception allows institutions to make digital reproductions for the preservation of works that are permanently in their collections. Cultural heritage institutions can therefore digitize works that are still under copyright; however, they are not allowed to provide access to these digital assets. The second exception allows them to make use of their assets for the purpose of text and data mining; it thus enables the application of machine learning procedures and the development of artificial intelligence applications. Cultural heritage institutions can therefore develop such secondary products on their own, be they machine translation models or large language models created out of

the available massive textual databases. The European Directive on the Legal Protection of Databases (European Commission, 1996) provides the legal basis for cultural heritage institutions to license their contents and thus to regulate access for their users as well as for private companies. Moreover, and according to the Data Governance Act (European Commission, 2020, currently in its approval phase), it is possible for institutions providing data sets and models to demand fees and to realize profits; with respect to fees, it is even possible to differentiate between small and medium-sized businesses (SMBs) and larger companies, such as the big tech companies. While digitization is mostly state-funded, the significant maintenance costs associated with the management of the digital assets—such as technical equipment, electricity, and human resources costs—can therefore be covered, at least partly, through such fees.

Such a juxtaposition of cultural heritage institutions with big tech companies highlights the changing functions of cultural institutions in the 21st century—they have gone from preserving physical assets to establishing and administering outputs of machine learning along with providing quality assurance for these products, for example, by preparing "Model Cards for Model Reporting" (Mitchell et al., 2019). Moreover, the aim of cultural heritage institutions to develop secondary products on their own opens up the possibility of strengthening their bonds with registered users, for example, by including them in the establishment of machine learning procedures. Users may become engaged in crowdsourcing activities, such as annotating images or collectively putting captions on them, labelling data, or enriching metadata. Such approaches foster the traditional idea of the commons, where members of the community are obliged to fulfil specific duties, they strengthen social sustainability and the maintenance of resources, and thus contribute to the sovereign and self-organized management of the commons.

3 CONCLUSION

Both challenges for the sovereign management of digital assets identified here point in the same direction, which may be described as a movement from the open to the closed. It is generally possible to maintain the commons that enables digital sovereignty and self-organization of cultural heritage institutions within current legal regimes. But whereas digitization was begun before the turn of the millennium following the ideal of providing full open access to digital assets, the protection granted by copyright and intellectual property rights can only be lifted by the provision of closed spaces in which users can access the digital material under certain conditions. The above-described developments have consequences in the form of reviving a classical and historical conception of commons as closed spaces with resources to which only an elite has access. In these elitist communities, self-regulation, trust in the normative framework, and the importance of

obligations serve to maintain the resources, while non-members of the community must ask for access to licensed content and have to pay fees. The downside clearly lies in a compartmentalization of the internet and in a disbanding of the idea of open access for everyone.

4 ACKNOWLEDGMENTS

This contribution is the result of a research project conducted in 2021/2022 at the Center for Advanced Internet Studies (CAIS) in Bochum. The author would like to express his sincere gratitude to the CAIS and the Ministerium für Kultur und Wissenschaft NRW, which funded this research project. He is particularly indebted to the wonderful team at CAIS College; the law scholars Brett Frischmann, Michael Madison, Katherine Strandburg, and Madelyn Rose Sanfilippo, who developed the Knowledge Commons Framework used in the analysis and participated in a workshop conducted during the fellowship; to all the interviewees who took part in the study; and to the fellow researchers at CAIS and elsewhere with whom he exchanged on this subject.

5 REFERENCES

1. Bender, E., Gebru, T., McMillan-Major, A. & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots. Can Language Models be Too Big? Conference on Fairness, Accountability, and Transparency (FAccT '21), March 3-10, Canada. ACM, New York, NY, USA, pp. 610 –623. https://doi.org/10.1145/3442188.3445922.
2. Boyle, J. (2008). The Public Domain. Enclosing the Commons of the Mind. New Haven / London: Yale University Press.
3. European Commission (1996). Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases. Brussels: European Commission. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31996L0009:EN:HTML.
4. European Commission (2019). Directive on Copyright in the Digital Single Market – Directive (EU) 2019/790. Brussels: European Commission. http://www.europarl.europa.eu/doceo/document/A-8-2018-0245-AM-271-271_EN.pdf.
5. European Commission (2020). Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act) – COM/2020/767. Brussels: European Commission. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0767&from=EN.
6. Frischmann, B. M., Madison, M. J., Strandburg, K. J. (2014). Governing Knowledge Commons. In B.M. Frischmann, M. J. Madison & K. J. Strandburg (Eds.), Governing Knowledge Commons. Oxford: Oxford University Press, pp. 1–43.
7. German Commission for UNESCO (2021). Bundesweites Verzeichnis Immaterielles Kulturerbe: Demoszene – Kultur der digitalen Echtzeit-Animationen. https://www.unesco.de/kultur-und-natur/immaterielles-kulturerbe/immaterielles-kulturerbe-deutschland/demoszene.
8. Haux, D. H. (2021). Die digitale Allmende. Zur Frage des nachhaltigen Umgangs mit Kultur im digitalen Lebensraum. Zürich/St. Gallen: Dike Verlag. https://doi.org/10.3256/978-3-03929-012-3.
9. Jo, E. S. & Gebru, T. (2020). Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning. FAT*'20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, pp. 306-316. https://doi.org/10.1145/3351095.3372829.
10. Lehmann, J. (2022). The Tragedy of the Cultural Commons. Research Report and Data Publication. https://doi.org/10.5281/zenodo.6513596.
11. Lenski, S.-C. (2013). Öffentliches Kulturrecht. Materielle und immaterielle Kulturwerke zwischen Schutz, Förderung und Wertschöpfung (= Ius Publicum, 220). Tübingen: Mohr Siebeck.
12. Mitchell, M. et al. (2019). Model Cards for Model Reporting. FAT*'19: Proceedings of the Conference on Fairness, Accountability, and Transparency, pp. 220–229. https://doi.org/10.1145/3287560.3287596.
13. Yakowitz Bambauer, J. R. (2011). Tragedy of the Data Commons. Harvard Journal of Law and Technology 25:1, pp. 1–67. https://doi.org/10.2139/ssrn.1789749.

Proceedings of the Weizenbaum Conference 2022:
Practicing Sovereignty. Interventions for Open Digital Futures

OPENING SCHOOLS TO STUDENTS' INFORMAL DIGITAL KNOWLEDGE TO ENABLE THE EMANCIPATORY EMPLOYMENT OF DIGITAL MEDIA

Heinz, Jana
German Youth Institute
Munich, Germany
heinz@dji.de

KEYWORDS

digitality; digitization; educational equity, equality; informal knowledge; elementary schools; digital divide; digital education

DOI: 10.34669/wi.cp/4.16

ABSTRACT

While classes become more heterogeneous and children grow up as digital natives, instruction is still characterized by an emphasis on middle-class children and analogue media. Moreover, national and international comparative studies have repeatedly shown that Germany in OECD comparisons often ranks last in terms of the level of digital learning opportunities in schools. A gap exists between children's lifeworld experiences and informal learning processes in a digital world on the one hand and digital learning opportunities at school on the other. Thus, schools do not offer content and digital infrastructure that links to students' informal digital knowledge. Therefore, there is a need to discuss how schools can integrate the emancipatory power of digitalization.

1 THE DIGITAL DIVIDE

Well-known problems of educational institutions remain in a digital world. Students from low socio-economic backgrounds tend to not benefit from educational opportunities at school as much as students from higher backgrounds (Bourdieu & Passeron, 1977). These problems have been well documented in sociology since the expansion of education in the 1960s and have also been discussed widely by the general public since the first PISA publications. These injustices seem to have been further intensified by digitalisation (Ma, 2021). Thus, a digital divide has become apparent (Robinson et al., 2015)—again to the disadvantage of children and young people from low socio-economic backgrounds. Although they are familiar with the use of digital media, they still have low digital skills (Ghobadi & Ghobadi, 2013). Studies show that this is not so much due to digital family equipment (first-level divide). Rather, inequality is reinforced by differences in media use (second- and third-level divide) (Scheerder et al., 2017) and, above all, by differences between their informal digital knowledge and school requirements (Heinz, 2016).

While children from socially weaker milieus seem to use digital media in their free time more often than children from higher social milieus, this is not automatically accompanied by a learning advantage. The acquisition of digital competences depends not only on the frequency of use, but also on skills such as reading skills and dealing with complex information. In addition, digital learning opportunities are often quite challenging. For example, they require children to learn independently, yet, children with learning difficulties sometimes need additional support. At the same time, it is evident that children from socially weaker backgrounds are familiar with digital media and thus highly motivated to work with it in schools.

However, while classes become more heterogeneous and children grow up as digital natives, instruction is still characterized by an emphasis on middle-class children and analogue media. Moreover, national and international comparative studies have repeatedly shown that Germany as a whole often ranks last in terms of the level of digital learning opportunities in schools. Thus, schools do not offer learning and teaching that links to students' informal digital knowledge. Therefore, there is a need to clarify how schools can integrate the emancipatory power of digitalization.

2 ANALOGUE SCHOOLS (LARGELY) IN A DIGITAL WORLD

A look at practice and research in education reveals that digital media in schools is seen more as a tool for optimizing learning processes, and rarely as part of a changed, digital world (Krommer et al., 2019). Only in a few schools can pedagogical concepts and school routines be identified that exploit the potential of digital media for active and creative use in education: "When digital media are used

in lessons in German schools today, they are generally used for presentations, research on the internet or worldwide web, or reading in PDFs. Two aspects in particular stand out: Firstly, teaching with digital media appears to be primarily receptive and not very active. Secondly, it is apparent [...] that in very few schools teaching goes beyond the implementation stage of substitution, i.e. the replacement of analogue media with their digital equivalent." (Knaus, 2017, p. 58, author tr.). Uta Hauck-Thum and Noller (2021) argue even under the conditions of digitalization and digitality, teaching and learning processes are primarily oriented towards the print-media book culture. If digital media is employed at all in schools, then as tools to replace analogue media but they neither influence teaching structures nor encourage children to participate.

Obviously, a gap exists between children's lifeworld experiences and informal learning processes in a digital world on the one hand and digital learning opportunities at school on the other. Today, children and young people grow up in a world with digital media as a matter of course and are therefore regarded as "digital natives." However, this picture does not stand up to closer analysis when looking at the digital competences of children and young people: These competences are low level and mainly comprise user knowledge, i.e. simple surfing the internet, researching terms or clicking/opening apps (Aesaert et al., 2013). Without didactic and pedagogical support, students apply this knowledge superficially. Neither transferable action knowledge for the confident use of application software nor an understanding of the safe handling of data is built up by students on their own (Litt, 2013).

However, the term digital natives does make sense when one looks at the extent to which children are now growing up with digital media, as shown by the study "DIVSI U9 Study - Children in the Digital World" (DIVSIO, 2015, p. 6), which examined the media use of children aged 3 to 8. The authors conclude that it is no longer a question of whether children of this age should already use digital media. Rather, children have long been moving autonomously in a digital world and have a great interest in digital media. "Around 1.2 million 3- to 8-year-olds are regularly online. Children who cannot yet read and write recognize corresponding symbols that enable them to call up web offers." The KIM study, which examines the media use of 6- to 13-year- olds in Germany, draws similar conclusions. It reports, "42 percent of girls and boys use a mobile phone or smartphone every day, and at 35 percent, one in three listens to music almost daily. A good quarter of the children use the internet daily." (Feierabend et al., 2017, p. 10, author's translation). Digital media is thus an integral part of the lifeworld of children and adolescents and thus a significant influencing factor in their primary socialization. Children learn to use digital media as a cultural technique—such as reading and writing later on—in often informal learning processes in everyday life.

If children already bring digital user knowledge with them to school, the question of whether children can or should already learn with digital media in schools is outdated. Rather, it is now a question of how school-based learning can be connected with children's various digital competences and strengthens them in the confident use of digital media. Moreover, knowing about the digital divide alone is an argument for the expansion of school curricula to include digital competences. Only in this way can schools fulfil their educational mandate to prepare pupils from *all* social backgrounds for future living and working environments, which will be shaped even more in the future by digital technologies.

3 SCHOOL STRUCTURES IMPEDING DIGITAL EDUCATION

Three main factors explain the gaps between students' informal digital knowledge and schools' focus on analogue teaching. These include, firstly, the typical discourses on digital forms of teaching and learning in schools in Germany, which are often limited to the vulnerability of young children in particular and thus overlook its potential. Moreover, the unclear and sometimes contradictory data on the effects of digital media on learning processes plays a particularly important role in understanding the hesitancy to open up forms of teaching and learning to digital changes. Secondly, binding guidelines for the implementation of school development concepts with a focus on digital teaching and learning have only been recently introduced. Thirdly, typical school functional logics have hindered the integration of social changes and thus of digitalization into school structures. These factors will be analyzed below.

3.1 DISCOURSES ABOUT DIGITALIZATION

A look at the social sciences reveals major differences between the definitions of digitalization in different disciplines. In media cultural studies and sociology, for example, the newly emerging communication technologies have been studied with regard to their social effects since the 1950s. Amitai Etzioni (1968), for example, asks how people can use them to authentically and actively shape their own society and where the dangers of being dominated by them lurk. Similarly, in his 1986 book The Postmodern Condition, Lyotard ([1982] 2002) explores how knowledge becomes integrated into social structures when it is no longer legitimized by metanarratives (such as beliefs in progress). In particular, under the conditions that all people are guaranteed access to knowledge, via online databases, he describes opportunities for a new scope for plurality of knowledge. As a prominent representative of media studies, McLuhan ([1964] 2008) in turn shows how technologies and electronic media change perception and culture globally. What these concepts have in common is that

digitalization is not limited to technical concerns. They instead show how individual preferences, technical and cultural processes influence each other.

Current theories in this tradition, such as Felix Stalder's “The cultural condition” (2017) or the concept of “post-digitality”, initially emphasize the self-evidence of digital worlds. “Being digital will be as normal as breathing air and drinking water. Only once digital devices don't work will we remember them.” This is how Nicholas Negroponte (1998) describes the ease with which we (will) have become accustomed to the digital infrastructure of our lives. Kim Cascone (2000) refers to this now invisible self-evidence of the former “digital revolution” in the economy, culture, and life of every individual in his use of the term post-digital, which has since found its way into recent works on digitalization or digitality. As in the first concepts (Etzioni, Lyotard, McLuhan) two contrasting digital futures are usually sketched (cf. e.g. Stalder 2017), one as utopian (freely accessible knowledge, technology and technologies of participation) and the other as dystopian (post-democratic world of surveillance and capitalist knowledge monopolies).

Digitalisation as a topic for education, again has a specific framing. In terms of time, three different phases can be distinguished, even if they overlap and are rather heuristic in nature: Initially, the critics of digital media dominated public perception—especially in the feature pages of major magazines (Büsching & Riedel, 2017). The scenario of digital dementia conjured up, for example, by Manfred Spitzer—brain researcher and critic of digital games and learning opportunities—is paradigmatic of this, arguing it threatens young people if parents do not protect them from digital media. These warnings still seem to dominate the attitudes of many parents, especially in the middle classes with their strong emphasis on education.

The subsequent phase focuses primarily on the “added value” (“Mehrwert”) of using digital media in classroom teaching and learning settings and emphasizes the “primacy of the pedagogical” (cf. critically Krommer, 2021). Additionally, scientific studies on the learning effects of digital media referred to in this context were highly contradictory, as shown, for example, by the results of the meta-study published in 2009 by the learning researcher John Hattie. In a systematic review of more than 800 studies on factors that positively influence learning outcomes, Hattie also examined computer-based teaching. According to Hattie, most of these forms of learning, such as internet exercises or simulated games, had little to no effect. Only interactive learning videos achieved a measurable positive learning effect (Hattie, 2009). In contrast, there are studies that focus on the effects of the targeted use of digital learning opportunities to assist children from socially disadvantaged backgrounds (Ma, 2021). Schachter and Booil (2016), for example, show how preschool teachers were able to significantly improve the mathematics skills of children with learning

difficulties from socially disadvantaged families in a short period of time through the use of special software learning programs.

Finally (as the third phase), the perception that digitalization is also fundamentally changing education itself, in the sense of a comprehensive cultural change (KMK 2021), is gaining ground. For example, in 2021, the Conference of German State Education Ministers (*Kultusministerkonferenz*, or KMK for short) published a supplement to the strategy "Education in a Digital World" from 2016, documenting this change. The newly published supplement "puts into perspective the path from 'teaching and learning with digital media and tools' to learning and teaching in a constantly changing digital reality, which becomes evident as a digital culture, particularly in cultural, social and professional modes of action, and in turn triggers the digitalization processes." (KMK, 2021, p. 3, author's translation)

These changes affect education, educational institutions and access to knowledge. Thus, the plurality of knowledge institutions is emerging, which include digital knowledge databases such as Wikipedia. Via digital devices, knowledge is decentralized and accessible to all: "The classrooms and lecture halls of yesteryear are dead, although you still find them everywhere and although society [...] still wants to impose them on us" writes Michel Serres (2015, p.38) to illustrate the extent of cultural-technological changes for each individual as well as educational institutions.

3.2 NO LONG-TERM FOCUS ON DIGITALIZATION IN SCHOOLS

State infrastructure was lacking for a long time. Only since 2016 have the German Government and the federal states made extensive financial resources available to create digital infrastructures (see, e.g., DigitalPakt Schule of the Federal Ministry of Education and Research). The disbursement of these funds is linked to the condition that the schools applying for this funding prepare school development plans focusing on digitalization. These plans must include descriptions of the planned integration of digital teaching and learning settings, what pedagogical concepts will be employed and how teachers will be trained. Furthermore, the teaching of digital competences has been anchored in the curricula, educational plans and framework curricula of the federal states since 2016. Accordingly, digital competences should be taught beginning in primary schools and continue, not as an additional subject, but as an integral part of all subjects (KMK, 2016).

With these federal and state digital packages and the inclusion of digital competences in the curricula, the educational policy framework for the comprehensive digitalization of schools has been set. Some principals had already tackled the digitization of their schools, but typically on their own and often with time-limited and project-based initiatives. Likewise, individual teachers have been using digital teaching and learning tools in their classes for a long time and have shared their

experiences on social networks. Well-known online-sites include Lehrer-Online (https://www.lehrer-online.de/) as well as Edupunks and Twitterlehrerzimmer (formerly EdchatDE) on Twitter (#twlz #twitterlehrerzimmer).

Since comprehensive infrastructures were lacking for a long time, the existence of digital teaching and learning concepts in schools depended on the commitment of individual school principals. Accordingly, schools still differ with regard to their degree of digitalization. This heterogeneity was evident during the conversion to distance and hybrid teaching in response to the COVID-19 pandemic. Not surprisingly, those schools that had already integrated digital forms of teaching and learning before the restrictions came into effect in March 2020 had a distinct advantage (OECD, 2021). Here, students could be better served with learning opportunities, teachers felt less burdened by the change to distance or hybrid teaching and reported that they were able to prevent the exacerbation of educational inequalities related to the socio-economic background of the students. In contrast, most teachers and learners in those schools, which had no digitalization strategy—especially in primary schools—, were overwhelmed by the abrupt switch to distance and hybrid teaching. Here, compared to regular school attendance, (digital) instruction was reduced and focused predominantly on the core subjects. Accordingly, many parents wished for more intensive contact and more advice on how to support their children. The lack of technology did not seem to be the main reason for limited teaching. Almost all households in Germany had internet-enabled devices (Porsch & Porsch, 2020).

3.3 SCHOOL LOGICS

Thirdly, schools are defined by a specific organizational logic that shapes their ability to integrate digitalization. Helmut Fend describes challenges that school actors face when they seek to integrate societal changes into schools. He speaks of a *re-contextualization* that becomes necessary (Fend, 2006). Similar challenges have also become evident with regard to digitalization. School leaders are required to integrate digital learning environments into their schools, yet they have to link these to specific conditions for action, such as the school infrastructure, the expectations of the teachers and the parents as well as the needs of individual children. This is particularly difficult when schools are overburdened by reform projects (inclusion, all-day schooling, increasing heterogeneity of pupils) taking place at the same time, the innovations are highly complex (maintenance, disposal of digital devices, uncertainties regarding applicable data protection regulations) and cannot fully be linked to internal school norms and established practices. In particular, teachers must be convinced of the innovations' benefits and be able to work with them; accordingly, training in initial and further education is necessary if digitalization is to be a permanent feature of teaching practice (Heinz, 2018).

4 CONCLUSION: KEEPING THE FOCUS ON THE EMANCIPATORY POTENTIAL OF DIGITAL MEDIA

Digital media are an integral part of children's and young people's lives and thus a relevant influence on their primary socialization at home. Moreover, with the onset of primary school, media ownership and consumption grow rapidly with each passing year. Accordingly, children grow into an independent use of the digital world, often before they learn to employ it more systematically. Children acquire the use of digital media as a cultural technique—like reading and writing later on—often in informal learning processes in everyday life.

However, in particular with regard to schools, digitalization is not limited to a plug-in-and play /learn of digital devices but requires the integration of socio-technical interdependencies that range from digital devices to children's hybrid prior knowledge, virtual worlds, data protection, and the economic interests of a digital capitalism. This places a variety of demands on schools. In addition, digitalization as a social change is accompanied by changes in the world of work and life, and new educational tasks arise in order to prepare children for sovereign participation in a thoroughly digitized world (such as the "new" competences of creativity, communication, collaboration, critical thinking, OECD, 2020). Thus, digitalization increases the number of objectives, such as imparting knowledge to children with different learning backgrounds, balancing out educational inequalities and, at the same time, allocating them to different educational paths.

Individual teachers and school principals alone cannot meet this multitude of demands. It requires their cooperation. Further, on the part of educational policy and administration, binding specifications in the form of reliable infrastructures (curricula, training, and further education etc.) as well as technical assistance in the procurement and maintenance of digital networks and devices is needed.

In view of these challenges, it is important to keep the focus on the emancipatory potential of digital media, which includes students' access to knowledge, the diverse previous experience of children with digital media and their creative use of it. However, this emancipatory power can only be unleashed if children are taught the competences to achieve digital sovereignty through schools.

5 REFERENCES

1. Aesaert, K., Vanderlinde, R., Tondeur, J., & van Braak, J. (2013). The content of educational technology curricula: a cross-curricular state of the art. Educational Technology Research and Development, 61(1), 131–151. https://doi.org/10.1007/s11423-012-9279-9
2. Bourdieu, P., & Passeron, J.-C. (1977). Reproduction: In education, society and culture (2. print). Sage studies in social and educational change: Vol. 5. Sage.
3. Büsching, U., & Riedel, R. (2017). BLIKK-Medien: Kinder und Jugendliche im Umgang mit elektronischen Medien. https://www.drogenbeauftragte.de/fileadmin/Dateien/5_Publikationen/Praevention/Berichte/Abschlussbericht_BLIKK_Medien.pdf
4. Cascone, K. (2000). The Aesthetics of Failure: "Post-Digital" Tendencies in Contemporary Computer Music. Computer Music Journal, 24(4), 12–18. https://ccrma.stanford.edu/~ananm/DAT330/CMJ24_4Cascone.pdfGesendet: Mittwoch, 07. Juli 2021 um 11:02 Uhr
5. Deutsches Institut für Vertrauen und Sicherheit im Internet (DIVSI) (Ed.). (2015). DIVSI U9-Studie: Kinder in der digitalen Welt. SINUS Institut, in Kooperation mit dem Erich Pommer Institut. https://www.divsi.de/wp-content/uploads/2014/02/DIVSI-U25-Studie.pdf
6. Etzioni, A. (1968). The active society.: A Theory of Societal and Political Processes. The free Press. Feierabend, S., Plankenhorn, T., & Rathgeb, T. (2017). KIM-Studie 2016. Kindheit, Internet, Medien. Basisstudie zum Medienumgang 6- bis 13-Jähriger in Deutschland. Medienpädagogischer Forschungsverbund Südwest. https://www.mpfs.de/fileadmin/files/Studien/KIM/2016/KIM_2016_Web-PDF.pdf
7. Fend, H. (Ed.). (2006). Neue Theorie der Schule. VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-90169-5
8. Ghobadi, S., & Ghobadi, Z. (2013). How access gaps interact and shape digital divide: a cognitive investigation. Behaviour & Information Technology, 34(4), 330–340. https://doi.org/10.1080/0144929X.2013.833650
9. Hattie, J. (2009). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. Routledge.
10. Hauck-Thum, U., & Noller, J. (Eds.). (2021). Digitalitätsforschung / Digitality Research. Was ist Digitalität? Philosophische und pädagogische Perspektiven. J.B. Metzler.
11. Heinz, J. (2016). Digital Skills and the Influence of Students' Socio-Economic Background. An Exploratory Study in German Elementary Schools. Italian Journal of Sociology of Education, 8(2), 186–212. https://doi.org/10.14658/pupj
12. Heinz, J. (2018). Zwischen Bereicherung und Belastung.: Einführung digitaler Medien in Grundschulen. Lernende Schule, 18, 42–45.
13. Knaus, T. (2017). Pädagogik des Digitalen. Phänomene – Potentiale – Perspektiven. In S. Eder, C. Mikat, & A. Tillmann (Eds.), Schriften zur Medienpädagogik: Vol. 53, Software takes command: Herausforderungen der „Datafizierung" für die Medienpädagogik in Theorie und Praxis (pp. 49–68). kopaed. https://www.pedocs.de/volltexte/2017/14797/pdf/Knaus_2017_Paedagogik_des_Digitalen.pdf

14. Krommer, A. (2021). Mediale Paradigmen, palliative Didaktik und die Kultur der Digitalität. J.B. Metzler, Berlin, Heidelberg. https://link.springer.com/content/pdf/10.1007%2F978-3-662- 62989-5_5.pdf

15. Krommer, A., Lindner, M., Mihajlović, D., Muuß-Merholz, J., & Wampfler, P. (2019). Routenplaner #digitaleBildung: Auf dem Weg zu zeitgemäßer Bildung : eine Orientierungshilfe im digitalen Wandel. Verlag ZLL21 e.V; Ciando. http://ebooks.ciando.com/book/index.cfm/bok_id/2767020

16. Kultusministerkonferenz. (2016). Bildung in der digitalen Welt: Strategie der Kultusministerkonferenz. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2016/2016_12_08- Bildung-in-der-digitalen-Welt.pdf

17. Kultusministerkonferenz. (2021). Lehren und Lernen in der digitalen Welt. Ergänzung zur Strategie der Kultusministerkonferenz „Bildung in der digitalen Welt“: (Beschluss der Kultusministerkonferenz vom 09.12.2021). https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2021/2021_12_09-Lehren-und-Lernen-Digi.pdf

18. Litt, E. (2013). Measuring users’ internet skills: A review of past assessments and a look toward the future. New Media & Society, 15(4), 612–630. https://doi.org/10.1177/1461444813475424

19. Lyotard, J.-F. (2002). The postmodern condition: A report on knowledge (13. print). Theory and history of literature: Vol. 10. University of Minnesota Press.

20. Ma, J. K.-H. (2021). The digital divide at school and at home: A comparison between schools by socioeconomic level across 47 countries - Josef Kuo-Hsun Ma, 2021. SAGE PublicationsSage UK: London, England. https://journals.sagepub.com/doi/10.1177/00207152211023540

21. McLuhan, M. (2008). Understanding media: The extensions of man (Reprinted.). Routledge classics. Routledge.

22. Negroponte, N. (1998, January 12). Beyond Digital. Wired, 12(6). https://www.wired.com/1998/12/negroponte-55/

23. OECD (Ed.) (2020). Framework for the Assessment of Creative Thinking in PISA 2021: Third Draft. OECD. (2021). The State of School Education: One Year into the COVID Pandemic. https://doi.org/10.1787/201dde84-en

24. Porsch, R., & Porsch, T. (2020). Fernunterricht als Ausnahmesituation. Befunde einer bundesweiten Befragung von Eltern mit Kindern in der Grundschule. In D. Fickermann & B. Edelstein (Eds.), Die Deutsche Schule Beiheft: Vol. 16, „Langsam vermisse ich die Schule …“. Schule während und nach der Corona-Pandemie (pp. 61–78). Waxmann. https://www.pedocs.de/volltexte/2020/20229/pdf/DDS_Beiheft_16_2020_Porsch_Porsch_Fernunterricht_als_Ausnahmesituation.pdf

25. Robinson, L., Cotten, S. R., Ono, H., Quan-Haase, A., Mesch, G., Chen, W., Schulz, J., Hale, T. M., & Stern, M. J. (2015). Digital inequalities and why they matter. Information, Communication & Society, 18(5), 569–582. https://doi.org/10.1080/1369118X.2015.1012532

26. Schacter, J., & Jo, B. (2016). Improving low-income preschoolers mathematics achievement with Math Shelf, a preschool tablet computer curriculum. Computers in Human Behavior, 55, 223– 229. https://doi.org/10.1016/j.chb.2015.09.013

27. Scheerder, A., van Deursen, A., & van Dijk, J. (2017). Determinants of Internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. Telematics and Informatics, 34(8), 1607–1624. https://doi.org/10.1016/j.tele.2017.07.007

28. Serres, M. (2015). Thumbelina: The culture and technology of millennials ((D. W. Smith, Trans.)). Rowman & Littlefield International.

Proceedings of the Weizenbaum Conference 2022:
Practicing Sovereignty. Interventions for Open Digital Futures

MINODU

FOSTERING LOCAL SUSTAINABLE DEVELOPMENT THROUGH TECHNOLOGY AND RESEARCH

Fröbel, Friederike
German Research Center for Artificial Intelligence (DFKI GmbH)
Berlin, Germany
friederike.froebel@dfki.de

Lange, Carina
University of Arts Berlin
Berlin, Germany
c.lange@udk-berlin.de

Joost, Gesche
German Research Center for Artificial Intelligence (DFKI GmbH)
Berlin, Germany
gesche.joost@dfki.de

KEYWORDS

participatory design; local community networks; co-creation; DIY networking; collective awareness; community learning; knowledge transfer; science communication; community wireless; ICT4D; climate change

DOI: 10.34669/wi.cp/4.17

ABSTRACT

Rapid climate change is exposing subsistence farmers to enormous challenges, especially in Sub-Saharan Africa. Several foreign aid programs have been set up to cope with these issues, many of which have focused on technical solutions. However, there seems to be a large gap between scientific research and the needs of local communities. Besides focusing on new ways to improve the resilience of local food production, there is also an urgent need to adapt available knowledge to the local context. Based on experiences from a project to co-create community networks in Togo in 2020, we aim to empower local stakeholders, including farmers and scientists, to adapt existing knowledge of sustainable crop farming to current practices. New modes of knowledge exchange can be established with the help of participatory design. These methods may help to foster a collective approach to learning that enables people to cope with global challenges on a local level, all while valuing the traditional practices of local farmers and enriching them with scientific knowledge.

1 INTRODUCTION

Climate change and the steady increase of the world's population already pose great challenges to sustainable land management and the conservation of natural resources and will certainly do so in the years to come. This particularly applies to the African continent, where the population is predicted to double in the next 30 years.

Farmers in Sub-Saharan Africa are facing enormous challenges when it comes to the staple crop farming that is needed to feed local populations. Several foreign aid programs have been set up to cope with this challenge. Many of them have a focus on technical solutions and look at issues like at irrigation methods and technologies. However, networking and knowledge exchange between sectors is barely encouraged, even though the adaptation of knowledge to the local context and the networking of actors along the value chain are key factors. They enable actors to anchor the needed knowledge on climate change and to foster durable behavioral change, eventually leading to agricultural practices that are able to cope with the challenges and feed populations.

While it is important to further boost research that helps to increase the resilience of local food production, there seems to be an urgent need to translate already available knowledge so that local communities can understand and use it. Social and technical measures that could be helpful for subsistence farmers in reducing their exposure to climate change already exist, but local communities are unable to access them. Our presumption is that these communities could be a decisive player in climate change adaptation if they are enabled to take an active role in resource conservation, sustainable land management, and change management.

Research on sustainable land management in sub-Saharan Africa, for example by the West African Science Service Centre on Climate Change and Adapted Land Use (WASCAL), provides both raw and processed empirical data. However, they often do not consider the exchange of this knowledge between actors, the distribution of data, or the transfer and adaptation of new technologies, and practices to local groups with limited access to information and digital technologies. In addition, a multitude of languages, illiteracy, and a focus on oral transmission make it difficult for farming communities to gain access to the needed scientific knowledge. Often there is no awareness of the existence of this knowledge, and hence, it is not looked for.

With our approach, we aim to address the sustainable development goals (SDGs)[100] set out by the UN. We are particularly targeting SDG 2, "Zero Hunger," and SDG 11, "Sustainable Cities and Communities". Our aim is to empower local communities to co-create more sustainable and resilient farming methods through participatory methods of knowledge exchange. We will focus on the

[100] Sustainable Development, Department of Economic and Social Affairs of the United Nations. URL: sdgs.un.org

resources and needs of local farmers and explore how these can be addressed in participatory design sessions by using suitable digital and analogue approaches. By using co-creation and involving local experts, students, and community members, we aim to bridge the gaps between different languages, cultures, and skills.

In this project, our co-creators are subsistence farmers in the rural areas of northern Togo and students who grew up in those communities. Our intent is to encourage local stakeholders, including farmers, local decision makers, or scientific researchers, to engage in a knowledge exchange. Ultimately, we hope that agricultural practices that combine the best of both worlds—the latest scientific results as well as proven traditional practice developed. We are drawing on experiences from our prior project *Miadé* (2020), where we co-created community networks in Togo.

Figure 4. Graphical representation of the project's approach (Illustration by Mia Grote). (Figure 1. Project illustration.)

The aim of this project, called *Minodu*, is to close the implementation gap between scientific concepts and concrete, local challenges in land management (see Figure 1). At the same time, the intercultural project team from Germany and Togo intends to collectively design a space for knowledge exchange that allows actors to address global challenges on a local level and value the traditional practices of local farmers.

2 Background and Related Works

With our pilot project, Miadé (duration: April–December 2020)[101], funded by the Federal Ministry for Economic Cooperation and Development (BMZ) and co-facilitated by the Gesellschaft für Internationale Zusammenarbeit (GIZ), Germany, we developed an approach for co-designing

[101] Miadé – Local Community Networks in Togo. URL: togo.drlab.org

participation formats in Lomé, Togo. The project involved co-creating digital community networks using small and cheap single-board computers, with the aim to build self-contained local WiFi networks that enable knowledge creation and transfer, for example for educational contents, entrepreneurs, or an artist collective (see Fröbel and Lange et al., 2022). We learned that local stakeholders who are able to facilitate these community exchanges are an indispensable part of the project consortium. Thus, we established the role of "local leads," who were able to build bridges between researchers and the communities. Equipping them with design and moderation methodologies and co-designing methodologies that are adapted to the local context was a huge part of our project work.

We concluded that building trust and respect is a necessary precondition for establishing an intercultural dialogue and mutual learning circles. It takes a lot of time but is worth the effort. We see empathy for local needs and the cultural setting as a starting point for every collaborative practice. We aim to give as much space as possible to the Togolese members of our team, being conscious of the imbalance regarding both information and finances, as our project was framed and funded by a German public actor. We see ourselves as facilitators and are aware that our local partners possess the important knowledge and are key to achieving our objectives. Nonetheless, it is important to consider our past and coming projects in the context of post-colonialism and "white saviourism." Those are complex and important discussions and invitations to self-reflection that need to be continued at every stage of our project. Building on the knowledge and methodologies developed in the Miadé project, in the Minodu project, we will further elaborate on and carry forward our approach in more depth.

3 PROJECT SCOPE

The project name Minodu means "let's be together, let's work together" in Mina. Named after the town Elmina in Ghana, this dialect of Ewe is spoken in the south of Ghana, Benin, and Togo. French is the official language in Togo, as the country is a former French colony. However, the most common languages for everyday communication are Ewe and Kabye, as well as the various dialects of these languages. The diversity of languages presents, inter alia, the entry points for the Minodu project. How can the knowledge of a certain topic be captured, conditioned, leveraged, exchanged, and made available between different stakeholders, despite these linguistic challenges? How can people from small language communities in turn contribute to scientific research, for example in sharing their needs or good practices when it comes to crop resilience?

The focus of the four-year project is to create local and scalable participation formats for communities in vulnerable rural areas. The project will consist of three iterative workshop phases and

a concluding pilot phase. Each stage will last about a year and conclude with evaluations and deduced concepts for the following stage. We will collaborate with six communities of subsistence farmers in the rural area of Kara in the north of Togo. The core work will center on their specific needs, which will be collected and assessed in detail as a first step. The focus on their needs will increase motivation to be part of our project, as we are creating solutions of immediate interest to them.

Each community will collaborate with students and doctoral candidates from the Institute of Agricultural Professions (ISMA) of the University of Kara. The institute has experience in using a community outreach approach and is familiar with the surrounding communities and their needs. This provides a basis of trust. Local leads are a further part of the consortium, —they are experts, we have already worked with at the Miadé project (Fröbel & Lange et al., 2022). These local leads help to anchor the project implementation, contribute with their expertise, and guide the moderation of the workshops through their various stages according to different design methodologies.

Our task as researchers also consists in co-developing moderation guidelines with the local leads. This helps the respective facilitators to lead through the co-creation modules and to prioritize participant's input. We will create a "collective learning space" during our project. This includes the design, development, and implementation of a participatory tool consisting of low-cost hardware and free/libre/open-source software (FLOSS) for the creation and exchange of knowledge modules. Topics will include the adaptation of farming practices, the choice of local staple crops, or measures for disaster prevention in the light of climate change.

By "translating" existing studies into formats that can be applied directly in rural communities in the north of Togo, we will establish a mutual exchange of knowledge between the team of researchers and local stakeholders. We will combine analogue and digital technologies to create participation formats such as co-design sessions. Social media platforms, video tutorials, or podcasts will be potential levers of our communication strategy. The aim is to allow for exchange and participation in a customized manner that includes the use of different languages, video formats for illiterate people, and hands-on sessions for the application of knowledge. Content will be "translated" into these formats and ultimately be certified using Creative Commons in order to increase accessibility.

The results will then be evaluated, presented, and discussed in a concluding symposium involving all partners and stakeholders to enable scaling, independently of the project team.
Summing up, the three project goals are to:

- Co-design a participation format for knowledge exchange on regional land management with a focus on staple crop farming and with regards to climate change and climate adaptation.

- Encourage networking of different actors to bundle competencies, improve the use of resources, and anchor good practices.
- Co-design, provide, and further develop knowledge modules on topics such as climate change, sustainability, and land use with respect to issues such as water management and desertification, as well as associated technologies in relevant languages and for different levels of expertise.

3.1 METHODOLOGICAL APPROACH

The Minodu project thrives on participation and co-creation. It is based on and builds on our experiences with the Miadé project. This implies that the users, stakeholders, and different interest groups—the "communities" —are actively involved in the development process of the project and its objectives from the very beginning (Saad-Sulonen, 2018).

Furthermore, we draw on the concept of social learning spaces (SLSs). These thrive thanks to the engagement of participants who intend to share experiences and create change together (Wenger-Trayner & Wenger-Trayner, 2020, p. 19). In the process, learners develop agency, which Wenger-Trayner & Wenger-Trayner define as "the power to make a difference". SLSs thus go beyond the simple acquisition of knowledge; learners become directly involved in application and reflection, leading to a deeper and more sustainable learning experience. This in turn activates sustainable behavior change.

The aspect of community learning continues to be neglected in classical learning theories and in most (adult) education approaches. Learning has become more and more individualized in recent years (Azalde et al., 2019; Baker et al., 2019), which makes it even more difficult for learners in collectivist societies, such as those in Togo, to keep up. By collectivist societies, we mean less self-centered groups which emphasize common values above the needs of the individual. One example are e-learning courses, which are typically addressed to an individual learner. A farmer from the north of Togo, who may have difficulties in speaking French or might be illiterate, will struggle when attending such an online course. SLS allows us to investigate alternative scenarios: How can learners with different skillsets support each other in their learning journey and together pass certain milestones? Are there ways to intuitively link learners who are left behind to knowledge? "Social learning spaces are rarely designed directly because it is by definition something participants create together. Still, with the proper mindset, there is much you can do to help bring one into being," according to Wenger-Trayner & Wenger-Trayner (2020, p. 39). Our ambition is to create the space for such an SLS to emerge.

We follow the “design justice movement”, which considers the role of the designer as a facilitator rather than an expert in the design process (Costanza-Chock, 2020). We do not intend to impose a solution or apply a specific technology to the communities we are working with. Our aim is to co-design a format that allows the knowledge within the community to grow, using technology as a lever. Technology—in this case digital media formats and potentially digital community networks (see Miadé) —is only a means to enhance processes within the project; it is not an end in itself.

We thus aim to challenge the “traditional” approach of development work, which often exports knowledge or technology from the Global North to the South. In our design, we aim to lift our end-users—farmers who are marginalized in many aspects—up to a position as experts on their own experience and grant them an active part in the design process (Sanders and Stappers, 2008). Consequently, we do not see ourselves as “experts” but as designers who hold the space for members of the communities that are directly affected and are thus at the front and center of our design endeavor.

Togo was chosen as the locality of this project because our research team has been deeply linked with it for more than five years. Close connections to local NGOs and communities have been established, allowing collaboration on an equal basis and reducing power imbalances between North and South.

The farmers in the region of Kara are often disconnected from expert knowledge and digital information exchange: they are isolated in many senses and are facing disproportionate challenges linked to climate change at the same time. Following the “leave no one behind” (LNOB) principle of the United Nations (UN), their integration is key to achieving the United Nation’s Sustainable Development Goals (SDGs). As a project team, our ambition is to design SLSs by using our local network and contributing our design experiences.

Ideally, both the knowledge generated and the SLSs that are designed should further be applied in other communities who are marginalized and/or affected by climate change—both methodologically and in terms of their content. Through bringing this knowledge from South to North, we aim to question entrenched systems of power and dominance and twist the traditional North/South dynamic. We also intend to reflect the worldwide impact of climate change, the possibilities of local adaptions of expert knowledge, and the role of intercultural competences in development cooperation and research.

3.2 TECHNOLOGICAL APPROACH

We will work with "do-it-yourself" (DIY) participation approaches. These include, for example, local community networks (LCNs) that are independent of the internet and based on microservers. Those networks provide access to digital infrastructure, participation, and knowledge sharing in a dedicated local setting. FLOSS technology is used to strengthen, enhance, and expand already existing social connections. As LCNs are independent of the internet, they can also be used in the event of connection failures or natural disasters and independently of the radio cells provided by telecommunications service providers. We have proven the value of these LCNs in the Miadé field study and observed their value for local content production and exchange (Fröbel and Lange et al., 2022).

Arguments in favor of those networks include their capacity to stimulate communication within local communities and enable self-owned digital infrastructures. Installing LCNs offers communities the opportunity to open up to digital participation in general (Smyth & Helgason, 2019). However, accessibility is not only a question of the availability of technology but also depends on education, physical access, awareness, social structures, access to electricity, and technical know-how (Antoniadis, 2016, Unteidig et al., 2016). LCNs are only one possible approach to knowledge exchange in the project. We aim to develop other formats, considering intercultural aspects such as different languages, (digital) literacy, and social aspects. Therefore, we will facilitate co-creation workshops with local stakeholders as well as with researchers from the University of Kara to design appropriate media formats and evaluate their efficiency in an iterative process.

4 CONCLUSION AND OUTLOOK

The overall aim of the Minodu project is to bridge the gap between scientific research and local communities and provide analogue and digital participation formats to adapt available knowledge to the local context and open up access. Minodu encourages a networking of different actors to merge competencies, improve the use of resources, and reduce expenditures.

Our thematic focus lies on sustainable land use. We aim to strengthen capacities for sustainable local farming in times of climate change and develop a toolbox of methodologies and formats that can be transferred to other local contexts as well as to other topics.

We aim to encourage interdisciplinary knowledge exchange and raise awareness regarding intercultural competencies and the value of diverse research teams. A collaborative approach and a critical reflection on power structures within research teams is key to producing knowledge that is relevant locally and helps address the SDGs collaboratively. Spaces where local knowledge can be

expressed in the realm of scientific research are rare but can be designed. Here lies an underestimated lever to reaching the SDGs.

We are using the findings of our prior projects on community networks and adapting this knowledge to the specific context and topic. Our project team from Germany and Togo aims to collectively learn and design in a participatory manner, showing how to cope with global challenges on a local level, enhancing our intercultural competencies, and valuing the traditional practices of local farmers. We are therefore eagerly looking forward to the beginning of the Minodu project as a potential contribution to opening up new ways of participation and knowledge exchange.

5 ACKNOWLEDGMENTS

The project is funded by the German Federal Ministry of Education and Research (BMBF) under the category “Sustainable land management in sub-Saharan Africa: Improving livelihoods through on-the-ground research” as part of the BMBF’s ‘Research for Sustainability’ (FONA) strategy (FKZ 01LL2202A).

6 REFERENCES

1. Antoniadis, P. (2016). Local networks for local interactions: Four reasons why and a way forward. First Monday. https://firstmonday.org/ojs/index.php/fm/article/view/7123/5661 (Accessed: 7 July 2022).
2. Azalde, G. Jacob R. S. Malungo, Nchimunya Nkombo, Sarah Banda, Ravi Paul, Chibesa Musamba, & Arne H. Eide. (2019). Using the International Classification of Functioning, Disability and Health model in changing the discourse of disability to promote inclusive education in Zambia. In *The Routledge handbook of disability in Southern Africa* (p. 14). Routledge, Taylor & Francis Group.
3. Costanza-Chock, S. (2020) Design Justice - Community-Led Practices to Build the Worlds We Need. PubPub. Available at: https://design-justice.pubpub.org/pub/n8f4t51b/release/1 (Accessed: 3 May 2021).
4. Fröbel, F., Lange, C., Mandaroux, J., Ajavon, N., Wirth, A.-L., Tsamedi, V., Afanou, S., Foli-Bebe, O., Joost, G. (2022). Co-creation workshops for developing local community networks during a pandemic. Research for All, 6(1): 17. https://doi.org/10.14324/RFA.6.1.17.
5. Saad-Sulonen, J., Eriksson, E., Halskov, K., Karasti, H., & Vines, J. (2018). Unfolding participation over time: temporal lenses in participatory design. CoDesign, 14(1), 4-16.
6. Sanders, E., and Stappers, P. J. (2008). Co-creation and the new landscapes of design. 799–809. https://doi.org/10.1080/15710880701875068.
7. Shayne Baker, Malcolm Cathcart, & Neil Peach. (2019). Individualised learning approach (the three ‘p’s’) for a small to medium enterprise through work-based learning. 22nd Australian Vocational Education and Training Research Association Annual Conference. https://avetra.org.au/data/AVETRA_Final_2019_Baker_et_al.pdf.
8. Smyth, M., & Helgason, I. (2019). DIY community WiFi networks: Insights on participatory design. Conference on Human Factors in Computing Systems - Proceedings, June 2020. https://doi.org/10.1145/3290607.3313073.
9. Unteidig, A., Dominguez-Cobreros, B., Calderon-Lünig, E., Heilgemeir, A., Clausen, M. and Davies, G. (2016). D2. 1 Design, progress and evaluation of the Prinzessinnengarten pilot (version 1). http://www.mazizone.eu/wp-content/uploads/2018/07/MAZI_D2_2_final.pdf (Accessed: 7 July 2022).
10. Wenger-Trayner, B. and Wenger-Trayner, E. (2020). Learning to Make a Difference: Value Creation in Social Learning Spaces. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108677431.

Proceedings of the Weizenbaum Conference 2022:
Practicing Sovereignty. Interventions for Open Digital Futures

ARTISTIC INTERVENTIONS IN THE ICT INDUSTRIES

LEGITIMATE CRITICAL PRACTICE OR EMPTY GESTURES IN THE CONTEMPORARY DIGITAL AGE?

McDermott, Fiona
Trinity College Dublin
Dublin, Ireland
fiona.mcdermott@tcd.ie

Šiljak, Harun
Trinity College Dublin
Dublin, Ireland
harun.siljak@tcd.ie

KEYWORDS

digital transformation; art and technology; artistic practice; critical practice

DOI: 10.34669/wi.cp/4.18

1 INTRODUCTION

Much of the digital transformation is driven by commercial and organizational interests and is subject to the influence of major information and communication technologies (ICT) companies, software and platform providers, and technical research institutions. Within these contexts, there has been a marked increase in the incorporation of artistic and creative practices into the development of digital technologies and applications. From artist in residency programs in private technology companies to creative collaborations in academic research labs, the development and design of ICT technologies are increasingly subject to the input of artistic and creative practitioners.

More broadly, the recent uptick of arts-in-technology programs has been driven by the increasing urgency of issues such as algorithmic bias, data privacy, disinformation as well as larger movements towards addressing the wider ethical, socio-political and environmental implications of the digital transformation. At the same time, among enterprise and technology development leaders there seems to be a growing recognition that artists and creative strategies are beneficial in terms of critiquing underlying presumptions about the relative value of outputs and diversifying future project developments. Accordingly, perspectives on these artistic engagements in industry run the gamut from deeming them compromised and complicit (Wilk, 2016) to creating direct, meaningful engagement and critical participation with those that drive technological development in the digital transformation (Fraaije et al., 2022). There are a variety of strategic goals for industry to initiate these engagements, ranging from widening the perspectives of engineers to implicitly addressing social, economic, or political issues to using artists to leverage "innovation" and improved public relations (McDermott and Fieseler, 2021).

But the collaborative relationships between the ICT industry and artists necessitate further questions of interpretation and evaluation of the practices, values, and outcomes created. For example, how do artists engaged with ICT companies critique the industry while maintaining relationships for funding? Does the work of artists legitimize corporate decisions? In what different ways are value attached to artistic practice in these entrepreneurial/corporate contexts and how does these differ for the different parties involved? Does the undervaluation of artistic work persist in industry through "opportunities" like residencies? Can the participation of artists in ICT industrial research and development domains ever truly challenge the multifaceted problems of contemporary technologies and contribute to alternative technological futures?

2 ARTISTIC PRACTICE AND ICT RESEARCH

Drawing on the case of *the Department of Ultimology (DoU)*, an ongoing artistic project that has been situated in an ICT research and development setting, this enquiry explores how the arts can have a valuable impact at the foundational level of the digital transformation through its diffusion with the ICT industry and research. The *DoU* project was initiated in the context of an artistic residency at the CONNECT Centre for Future Networks in Trinity College Dublin. The project uses artistic methods to promote the hypothetical concept of *ultimology*, which refers to the study of endings. By paying close attention in the present to entities that are vulnerable or at risk to ending in different contexts, the DoU sets out to explore themes and questions around endings and imaginary possibilities.

One example of their artistic approach was the workshop as part of the Weizenbaum Conference 2022, whereby the *DoU* invited participants to work with and learn from turf—a highly contested material composed of deposited wetland vegetation, which, when extracted from bog environments in Ireland, dried, and compressed, can act as a fuel source. Turf is an entity at the intersection of contemporary conversations around environmental concerns, climate justice, value, tradition, natural capital, and energy usage in an Irish context, where technology and data centers are making rapidly increasing demands on the national grid. In the context of this workshop, turf is offered as an ultimological material; its sale and distribution is increasingly likely to be made illegal in the near future. During the workshop, participants were invited to encounter turf—to handle and transform it, while discussing its value as a contested and vulnerable material. Through tactile handling and transformation of this material, the workshop participants generated a discussion on the disputed and nuanced issues of value and endings.

3 CONCLUSION

Art can be used to express and represent imaginaries, ideas and phenomena previously unseen, leading to open discussion of things we may not have dared to think or express, and serve towards the concrete formulation of questions that can help move society forward. The work of the artists in this workshop permits imaginative experimentation that has the potential to forge new, unexpected, and alternative ways of thinking that disturb taken-for-granted and unreflective knowledge by initiating novel and unforeseen connections. In turn, by disturbing unreflective knowledge, imaginative experimentation, such as the experiments put forth by the *DoU,* give rise to deeper ethical considerations and fresh perspectives on technologies and innovation. Being unrestraint by the logics of established methodological procedures in engineering and science, the work of the *DoU*

demonstrates how art's open experimental methods permit a freer and less constrained form of inquiry that validates provocation, nonconventional methods, and speculation.

In conclusion, this research explores how the participation of artists in research and industry domains might challenge the multifaceted problems of contemporary technologies and contribute to alternative technological futures. While everyday understandings of sociotechnical relations often appear locked within a narrative paradigm that the current projected future is inevitable and unchangeable, artistic practice can operate as a means of opening up tacit knowledge and giving time and space to imagine and voice alternative ideas and theories that are otherwise unheard. Nevertheless, further questions remain in terms of preventing the instrumentalization of artistic practice in ICT industry and research contexts as well as the long-term commitments of ICT industry and research organizations to engage with artists over sustained periods of time.

4 ACKNOWLEDGMENTS

This research is part of the EU Horizon 2020 project Artsformation. The project explores the intersection between arts, society and technology and aims to understand, analyze, and promote the ways in which the arts can reinforce the social, cultural, economic, and political benefits of the digital transformation. The authors wish the thank and acknowledge the input of the artists and workshop participants in this research.

5 REFERENCES

1. Fraaije, A., van der Meij, M. G., Kupper, F. and Broerse, J. E. W. (2022). Art for Public Engagement on Emerging and Controversial Technologies: A Literature Review. Public Understanding of Science, 1– 17.
2. McDermott, F. and Fieseler, C. (2021). Mapping of Arts Integration within Enterprise. Report of the EU H2020, Artsformation Report Series
3. Wilk, E. (2016). The Artist-in-Consultance: Welcome to the New Management. E-Flux Journal. Online at: http://www.e-flux.com/journal/the-artist-in-consultance-welcome-to-the-newmanagement/

Proceedings of the Weizenbaum Conference 2022:
Practicing Sovereignty. Interventions for Open Digital Futures

MAKING ARGUMENTS WITH DATA

Savic, Selena
FHNW Academy of Art and Design
Basel, Switzerland
selena.savic@fhnw.ch

Martins, Yann Patrick
FHNW Academy of Art and Design
Basel, Switzerland
yannpatrick.martins@fhnw.ch

KEYWORDS

visual data study, situated knowledge, data observatories, machine learning, correlationism, critique from within

DOI: 10.34669/wi.cp/4.19

ABSTRACT

Whether we are discussing measures in order to “flatten the curve” in a pandemic or what to wear given the most recent weather forecast, we base arguments on patterns observed in data. This article presents an approach to practicing ethics when working with large datasets and designing data representations. We programmed and used web-based interfaces to sort, organize, and explore a community-run archive of radio signals. Inspired by feminist critique of technoscience and recent problematizations of digital literacy, we argue that one can navigate machine learning models in a multi-narrative manner. We hold that the main challenge to sovereignty comes from lingering forms of colonialism and extractive relationships that easily move in and out of the digital domain. Countering both narratives of techno-optimism and the universalizing critique of technology, we discuss an approach to data and networks that enables a situated critique of datafication and correlationism from within.

1 INTRODUCTION

The outputs of machine learning algorithms trained on large datasets (often referred to as "big data") play an increasingly important role in decisions that concern personal as well as global, socio-political and economic choices. The patterns and trends observed in machine learning models are taken as sources of truth and reason in recruitment and admission processes. They guide policymakers in deciding on measures to take in the pandemic, and they help individuals decide whether to purchase an item or not. Nevertheless, we have very limited access to examine the datasets that inform such decisions. Tools and frameworks such as Google's Colab[102], OpenAI's GPT3,[103] or design-specialized RunwayML[104], come with pretrained models and assumptions of correlation between data points. Professionals outside of computer science field work with these frameworks to quickly prototype experimental and innovative projects that are informed by machine learning (ML) and artificial intelligence (AI) tools.

Artistic practice has repeatedly demonstrated how hard it is to counter the assumptions and biases that permeate through automated training processes on datasets. For example, in a recent experimental theatre piece by artist Simon Senn and developer Tammara Leites titled *dSimon*,[105] an artificial personality was performed as a conversant, artistic advisor and as a stand-in for Elon Musk and Simon Senn himself. The dramatic unfolding of inappropriate behavior by the *dSimon* conversation agent, trained on Simon Senn's personal data using the GPT-3 artificial intelligence engine, engaged the audience as witnesses to the bizarre and unsettling propositions. The imaginary of neutrality in vast collections of internet-based text is quickly dispelled, revealing the inherent sociality of anyone's or anything's ability to understand and compose language.

This article combines the technical and artistic perspectives on the bias, and other forms of structural inequality in applications of machine learning models, informed by critical data studies and the critique of contemporary aspirations to objectivity in machine learning applications. The central argument engages the critique of scientific aspiration to universal objectivity most notably addressed by Donna Haraway (1988, 2016) and focuses on the critique of contemporary aspirations to objectivity in working with data, particularly in terms of information representation. This points to the need to envision different ways of working with datasets and machine learning models: These

---

[102] Colaboratory: browser-based machine learning environment, funded by Google; visit https://colab.research.google.com/ [accessed 15 February 2022].
[103] GPT-3 neural network machine, funded by Microsoft and Elon Musk; https://openai.com/ [accessed 15 February 2022].
[104] Runway ML, machine learning platform for visual tasks; https://runwayml.com/about/ [accessed 15 February 2022]
[105] More information on the performance of the *dSimon* theatre piece in Vidy theatre in Lausanne, in December 2021 is available at: https://vidy.ch/en/dsimon-0 [accessed 15 February 2022].

should entail enabling people to formulate arguments based on relations they actively discover in the data and trained models. A machine learning model is the output of a machine learning algorithm run on a specific dataset, which establishes data structures and relationships that can be applied to further datasets to infer similarities and predictions. Because such models are dependent on the training process and datasets, they tend to re-encode pre-existing determinisms and beliefs. In her work on race and technology, sociologist Ruha Benjamin identified this as engineered inequity and default discrimination (Benjamin, 2019). Furthermore, as computer scientist Cathy O'Neil has observed (O'Neil, 2016), decisions to take correlations between data on, for example, employment histories and addresses at face value, is at the root of the discriminatory operations of algorithms. In order to change this, we argue that we have to start from the dataset and reimagine the expectations of truth and reason from training processes and trained models.

Seeing things in data has historically been of interest in art and in engineering. The aspiration to make visible and public that which is measured and documented can be traced back to the 19th century, when it was first used to denote making visible the information that was not actually present at sight.[106] Data representation is rendered ever more accessible and efficient with the use of information technologies, and with this grows the responsibility to maintain the specificity and situatedness of the assumptions and inferences one makes when working with datasets. We believe that one can learn to do this with a carefully crafted digital tool that makes it possible to navigate datasets in experimental, non-essentialist ways, in order to practice a critique of datafication and correlationism from within. Based on the experiences with SNSF-funded research project *Radio Explorations*, discussed in more detail below, we expect that this way of working with data can result in meaningful arguments, engagements, and stories.

2 Resisting Colonial Relations in Machine Learning Processes

The current landscape of machine learning [ML] includes a large number of tools that are becoming increasingly accessible for people without a computer science background or any coding skills. These tools run on remote servers, as the computing power they require to solve the complex statistical analysis cannot be achieved with regular laptop or PCs.

Tools such as Runway ML provide artists and designers with prebuilt models and a pay-as-you-go system to deploy heavy computing ML on remote servers. For creative practitioners, the software

[106] The nonavailability to sight is mentioned in the entry on Visualization at the Oxford English Dictionary https://www.oed.com/view/Entry/224009. For a good historical overview of cultural importance of data visualizations, see Orit Halpern's book *Beautiful Data* (Halpern, 2014).

offers access to several basic ML models: text generation, image synthesis, and object detection. The result displays predictions by their trained algorithms based on an input by the user but without showing either the data nor the process of the prediction. Such approaches could be called "arboreal," to borrow the term from Deleuze and Guattari's *rhizomatic theory* (1976). It preserves a tree-like hierarchical conception of knowledge and information with discreet categorization. This confronts the user with a tool that does not grant access to the underlying technology of ML such as statistical analysis, data clustering, and prediction. By refusing access, such tools reproduce a colonial-like relationship of entitlement: Resources, such as computational power and algorithms, are claimed by those who operate them in their best self-interest, simultaneously organizing and extracting the work of their nomadic[107] users.

Scholars in social studies of science and technology have addressed the problems that arise with the use of pretrained ML algorithms as decision-making and forecast tools. These models tend to reproduce biases encoded in the data they are trained on. Such biases have already made their presence felt in automated decision making, which tends to exhibit racial and gender preferences in which job candidates to select, who to admit to a college program, who to incarcerate or grant parole to, or whom to give loan approval.

To counter such biases, US-based artist and researcher Caroline Sinders has led many workshops to create feminist datasets[108]. Data collection informed by intersectional feminist practices aspires to mitigate the effect of biases in ML algorithms by critically engaging in the data collection process (Sinders, 2020). Sinders' workshops invited the public to explore the meaning of data and its use for protest and social justice. In a related gesture, Crag Dalton and Jim Thatcher called for counter data actions (Dalton & Thatcher, 2014). Dalton and Thatcher offered provocations to the regime of "big" data that recognized its situatedness and the risk of technological determinism and challenged the notion of data being "raw." While current software for ML algorithms often lacks access to the data they are built upon, critical approaches to data collection in academic settings, or workshops within festivals and seminars, promote a discursive approach to the topic but lack a more technical approach.

To work with data means to take a position and to formulate a clear goal. Even if correctly translated as "given" from the original Latin term, data is not simply given and is always collected with certain logics of measurement and observation. Data and analysis never speak for themselves, as anticolonial pollution scholar Max Liboiron has poignantly illustrated (Liboiron, 2021). The

[107] Nomadic is used here to stress the non-settled status of online platforms users, who come and go, register, and depart; at the same time, the problem of user uprootedness resonates with Rossi Braidotti's nomadic theory, which addresses nomadic subjects resisting "deterritorialization" in Deleuzian terms (Braidotti, 2011).

[108] For an overview on Caroline Sanders's work, see: https://carolinesinders.com/feminist-data-set/ [Accessed 15 February 2022].

presumption of unproblematic and unaccountable (a special way of saying objective) data collection reproduces colonial relations to resources and reality. Liboiron also emphasized the importance of the care for the subject of critique. We therefore search for ways to develop and work with a digital tool that encourage critical engagement with data, involve formulating the questions that one wants answered prior to observing patterns in data, and clearly expressing one's position in regard to the question. We developed practical approaches to dataset making and interpretations of machine learning models, starting from the aforementioned archive of radio signals. With this, we hope to contribute a clear example for working with large dataset and machine learning technologies in an informed way that promotes participation and intentionality.

3 DATA OBSERVATIONS, PROJECTIONS, AND COMPARISONS

What patterns can we observe in data with our eyes? Our eyes provide us with an embodied, finite point of view (Haraway, 1988). Such point of view embodies limitations that are interpreted as polluting or disqualifying bias from a universalist, objective position. But a universally objective position implies having a way of being everywhere equally, the so-called "god's trick," which carries with it a denial of responsibility, to paraphrase Haraway. Our work with datasets and data representation offers a refreshed reading of Haraway's insistence on the importance and persistence of vision, in the face of the visualization of digital data and matters of representing data objectively.

In the *Radio Explorations* project, we designed data observatories as intuitive tools for orienting and navigating. The principal aim was to develop and practice techniques for working with digital data in a way that is ethically sensitive to biases and universalism and that highlights material and symbolic connections with the world. We developed and used a tool that enabled comparisons between patterns in datasets. Such practices foster a unique relationship between the data (given), the method of comparison and the questions one brings to the data.

For example, we created a dataset from the digital archive of radio signals by focusing on specific aspects of these situated recordings of radio transmissions. We computed features such as noisiness or the probability of silence in samples of radio signals found in the database. We then compared the measurement of similarity across signals—as established by a machine learning algorithm called Self-Organizing Map[109]—to those in other, not directly related datasets, such as a Free Music Archive (FMA) dataset for music analysis. By looking at music and radio signals from a comparable point of abstraction, we created a shared landscape of properties whereby data is

[109] SOM is an unsupervised machine learning technique introduced in the 1980s by a Finnish computer scientist Teuvo Kohonen (Kohonen, 1982). It is known for its ability to classify data in an intuitive manner, emergent from the data.

organized according to the conditions of the comparison. In the process, it becomes important how radio signal samples are placed next to each other: A direct similarity between radio signals on the map should reflect their likeness in an aspect that is shared with audible information on music. These comparisons opened up new readings of relationships that can be established across datasets and that refuse to lend themselves to causal interpretations and superficial correlations. While certain signals are similar to other signals in terms of protocols or applications (military, satellite communication, etc.), the setup described here makes it possible to disregard the instrumental qualities of telecommunications and focus on the way digital data can be articulated in its own terms. This means that digital data regarding radio signals is comparable to data on musical genres and that a certain inherent property of data can emerge from the comparison. Radio is therefore not understood in terms of its capacity to transmit messages, which recalls the problematic assumption of access and use of electromagnetic waves as a resource, as opposed to the capacity to conceptualize radio signals in terms of the digital traces they leave and how they interact with recording equipment—which is a perspective we develop in this paper.

The visual aspect of comparison and navigation is important. For example, Figure 1 illustrates the organization of the two previously mentioned datasets—that is, radio signals juxtaposed over musical genres. The visual qualities of radio signal spectrograms facilitate taking a distance from an instrumental perspective on radio signals and their usual categorization according to application or frequency. Signals are represented here in terms of abstract visual patterns that preserve partial qualities related to these instrumental concerns. Visual interpretation goes both ways: It is helpful to compare signals but also to perceive how the tool itself operates and question whether the connections proposed by algorithms actually make sense.

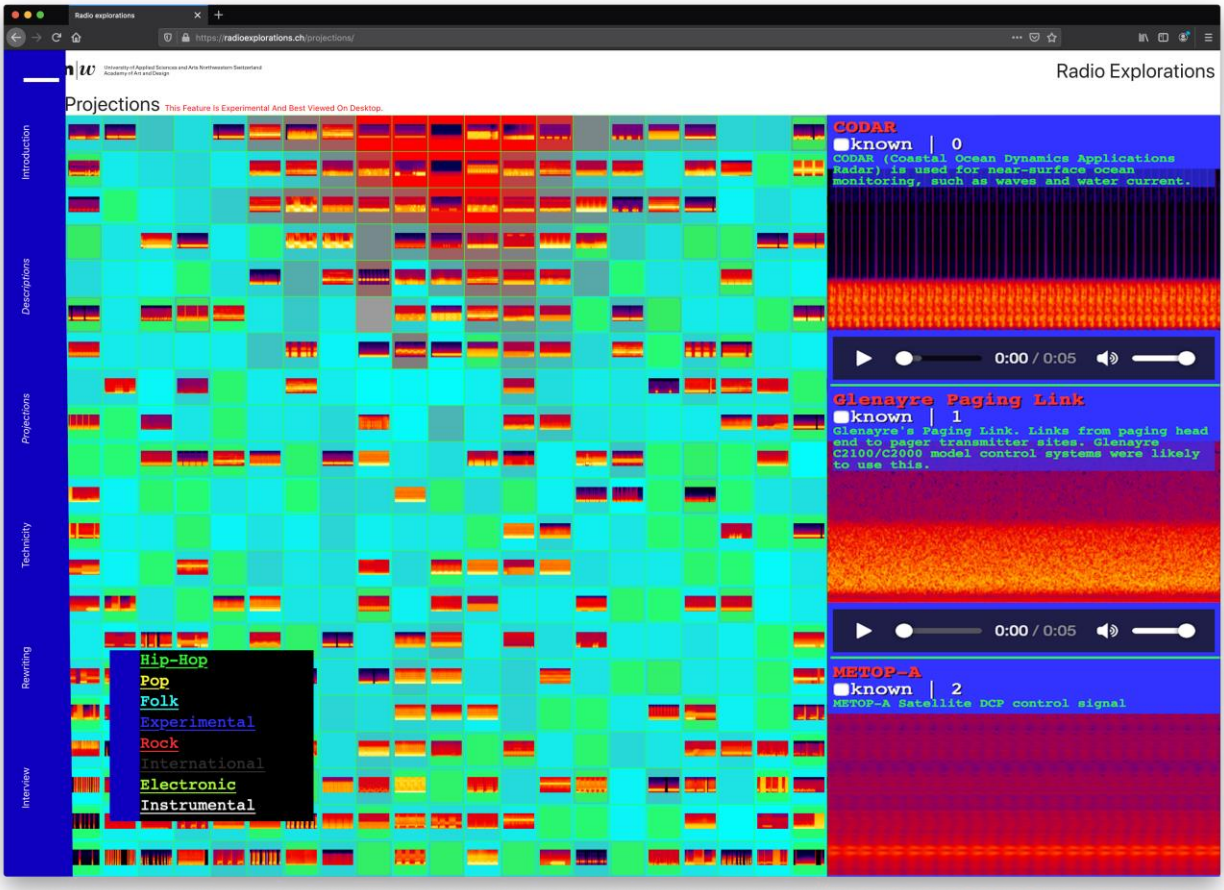


Figure 5. Radio Explorations. Signals are "projected" onto a preorganized map of musical samples, labelled according to the genre (overlay, bottom left). Each genre "highlights" some cells among which certain radio signals can be found. Highlighted here is the "Hip-Hop" genre.

We practiced working with different datasets and the idea of remaining open to the relationships in data, to the interpretability of statistics, and to data clusters. By selecting the data we worked with, and choosing a relationship we wanted to explore, we made it possible to search beyond correlations and establish meaningful comparisons across datasets so that people can make a visual/verbal argument that relates to their question and not to a "neutral" pattern in the data.

4 CONCLUSIONS

The recent rise of data driven technologies, like classifiers and recommender systems, has drawn attention to the problem of biases within data, and it has prompted vocal criticism of automated machine-learning-powered technologies. Nevertheless, such criticism often precludes alternative ways to use technologies that can be steered towards new modes of expression and argumentation. We hold that the main challenge for digital sovereignty and active participation in digital transformations actually comes from lingering forms of colonialism and extractive relationships that

easily move in and out of the digital domain. With this paper, we want to invite the reader to rethink ways of engaging with data so that people can take the space and structure to assert their own questions in relation to data.

We developed a technical framework that comprised a digital tool for data processing and analysis within (redacted) project and used it to explore multi-threaded narratives of music and telecommunication, of power and efficiency, encoded in datasets we worked with. We insisted on the visual organizing aspect of this practice. This is not meant as a call to improve ways of visualizing data but rather to innovate on ways to interpret and work with data using visual and other means to represent relationships always previously established in code (i.e., machine learning algorithms). Combining the concern for the importance and persistence of vision and its access to complex relations in the data, with the concern for digital sovereignty expressed as a resistance to colonial relations that haunt digital tools and knowledge of technical artefacts, we suggest paying attention to data in a carefully critical way.

5 ACKNOWLEDGMENTS

The *Radio Explorations* research project was generously supported by the SNSF-Spark funding grant number 190310. We are grateful to our research workshops guests: Carl Colena (SIGID wiki), Miro Roman (ETHZ), Simone Conforti (IRCAM), Sarah Grant (Kunsthochschule Kassel), and Roberto Bottazzi (The Bartlett) for their invaluable input. Special gratitude goes to Miro Roman for numerous informative discussions on working with SOM, as well as to Carl Colena for his support and discussions on radio signals beyond workshops and interviews.

6 REFERENCES

1. Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity.
2. Braidotti, R. (2011). *Nomadic Subjects: Embodiment and Sexual Difference in Contemporary Feminist Theory, Second Edition*. Columbia University Press.
3. Dalton, C., & Thatcher, J. (2014, May 12). What Does A Critical Data Studies Look Like, And Why Do We Care? *Society and Space*. https://www.societyandspace.org/articles/what-does-a-critical-data-studies-look-like-and-why-do-we-care
4. Deleuze, G., & Guattari, F. (1976). *Rhizome: Introduction*. Éditions de Minuit.
5. Halpern, O. (2014). *Beautiful data: A history of vision and reason since 1945*. Duke University Press.
6. Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, *14*(3), 575. https://doi.org/10.2307/3178066
7. Haraway, D. (2016). *Staying with the trouble: Making kin in the Chthulucene*. Duke University Press.
8. Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*(1), 59–69. https://doi.org/10.1007/BF00337288
9. Liboiron, M. (2021). *Pollution is colonialism*. Duke University Press.
10. O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy* (First edition). Crown.
11. Sinders, C. (2020, May 5). Rethinking Artificial Intelligence through Feminism. *CCCB LAB*. https://lab.cccb.org/en/rethinking-artificial-intelligence-through-feminism/

Proceedings of the Weizenbaum Conference 2022:
Practicing Sovereignty. Interventions for Open Digital Futures

HOW TO ENABLE SOVEREIGN HUMAN-AI INTERACTIONS AT WORK?

CONCEPTS OF GRASPABLE TESTBEDS EMPOWERING PEOPLE TO UNDERSTAND AND COMPETENTLY USE AI-SYSTEMS

Wienrich, Carolin
Julius-Maximilians-University
Wuerzburg, Germany
carolin.wienrich@uni-wuerzburg.de

Carolus, Astrid
Julius-Maximilians-University
Wuerzburg, Germany
astrid.carolus@uni-wuerzburg.de

Latoschik, Marc Erich
Julius-Maximilians-University
Wuerzburg, Germany
marc.latoschik@uni-wuerzburg.de

KEYWORDS

AI literacy; eXTended AI; human-centered AI; value-sensitive design; human-centered design

DOI: 10.34669/wi.cp/4.20

ABSTRACT

Artificial intelligence (AI) strategies are exhibiting a shift of perspectives, focusing more intensively on a more human-centric view. New conceptualizations of AI literacy (AIL) are being presented, summarizing the competencies human users need to successfully interact with AI-based systems. However, these conceptualizations lack practical relevance. In view of the rapid pace of technological development, this contribution addresses the urgent need to bridge the gap between theoretical concepts of AIL and practical requirements of working environments. It transfers current conceptualizations and new principles of a more human-centered perspective on AI into professional working environments. From a psychological perspective, the project focuses on emotional-motivational, eudaimonic, and social aspects. Methodologically, the project presented develops AI testbeds in virtual reality to realize literally graspable interactions with AI-based technologies in the actual work environment. Overall, the project aims to increase the competencies and the willingness to successfully master the challenges of the digitalized world of work.

1 INTRODUCTION

Digitalization is omnipresent in almost all areas of our lives. Developments in the field of machine learning and artificial intelligence (AI) are changing both our private and professional lives. Competencies and skills that allow for successful interactions with AI-based technologies are essential prerequisites for individuals to reap the societal benefits of digitalization (Law et al., 2018). The scope of public and scientific discussions of successful approaches of digitalization has long been limited to technical equipment and technical operating skills (Carolus & Wienrich, 2021; Wienrich & Carolus, 2021). Particularly in the context of work, the human has been unilaterally defined as *Homo economicus*, who regard technology as a tool to achieve goals, which are defined by the organizational framework and requirements (Carolus & Wienrich, 2021; Wienrich et al., 2022). More recent national and international AI strategies have argued for a more human-centric transformation process focusing on the human being and the individual's cognitive, emotional, and conative processes. For example, the German governmental AI strategy summarizes critical points of AI-related transformation processes as follows: "It is about individual freedom rights, autonomy, personal rights, the individual's freedom of choice. About hopes, fears, potentials, and expectations" (Bundesministerium für Bildung und Forschung, n.d.). In this context, the European Union has made considerable efforts to draft viable rules for the new world of AI. Within this framework, the EU High-Level Expert Group on AI established requirements such as: AI must be trustworthy (e.g., technically and socially robust) and respect human-centric approaches (e.g., respect human needs and diversity, avoid discrimination, be explainable). Thus, the previously merely technical focus was expanded to include an explicitly human-centric perspective. In particular, Article 14 of the AI regulation (Lexparency.org, n.d.) and the corresponding statements in the white AI paper (Madiega, 2022) emphasize that human needs should be taken into account in terms of both the design and the use of AI systems. Transparency and participation are regarded as essential elements to ensure trust in AI applications.

The present contribution emphasizes the urgency of translating these crucial but still rather theoretical demands into concrete professional practice. Furthermore, it bridges another gap: large parts of both public debates and governmental regulations refer to the business practices of only a few large high-tech companies. The reality of smaller and medium-sized companies and their employees is, however, neglected. The present project addresses these desiderata and focuses on people who work or will (soon) work with digital entities and AI-based systems—people who experience considerable digital inequalities and disparities. Workplaces are thereby considered to represent more to human beings than mere professional working spheres. Instead, workplaces are

essential social spaces that fulfill various human needs above and beyond professional efficiency and effectivity. Consequently, change processes at work affect employees at different levels, depending on the underlying basic human needs and interindividual differences: prior experience with technology and computer science, dystopian fears about AI or the individual's working environment and degree of participation, and digital disadvantages.

2 SOVEREIGNTY AND AI LITERACY AT WORK: RELATED WORK AND DESIDERATA

The concepts of media literacy and, more recently, digital literacy have been widely discussed as (rather) new cultural techniques that are an essential prerequisite for successful participation in our present and future digitized world. With the increasing importance of AI, these concepts of technology-related competencies need to be updated. The recently introduced concept of AI literacy (short: AIL) aims at understanding the competencies enabling people to successfully interact with AI-based technologies.

2.1 DESIDERATUM 1: THE NEED FOR VALID AND STRAIGHTFORWARD MEASURES OF AI LITERACY

The acquisition of digital (AI) literacy is crucial in order to competently harness future-oriented technologies. In a professional environment, there is a growing demand for new competency profiles that are increasingly characterized by associated AI-related innovations. Long and Magerko (2020) have analyzed competencies that enable the individual to comprehend, critically evaluate, and competently use AI technologies. The authors present a competency grid consisting of 17 AI-relevant skills, unified under the umbrella term *artificial intelligence literacy* (AIL; Long & Magerko, 2020). In an expert workshop, Wienrich et al. (2022), together with the AI Observatory of the German Federal Ministry of Labor and Social Affairs, expanded this framework. With an explicit focus on the work environment, their *Competence Behavioral Model of AI Literacy* (CBM-AIL) embeds individual and organizational potentials and barriers on the micro-, meso-, and macro-level. Besides these nascent conceptualizations, there are very limited measuring approaches that target general digital literacy (e.g., Jenkins et al., 2006; Ng, 2012; Ng et al., 2021; Porat et al., 2018) but not AIL (Wienrich & Carolus, 2021). Moreover, these measures are rather extensive questionnaires that are often unsuitable for practitioners or refer only to voice-based AI systems. Further studies analyzed perceptions and associations in the context of AI (e.g., European Commission, 2017; Hadan & Patil, 2020; Lau et al., 2018; Kelley et al., 2019; Zeng et al., 2017; Zhang & Dafoe, 2019). However, these studies often used measures consisting of only a single item, resulting in rather vague latent variables and limited validity. In summary, existing conceptualizations of AI-related competencies lack

practical and professional relevance. Additionally, the quality of existing measures is limited in terms of the scientific, practical, and work-related criteria.

2.2 Desideratum 2: The Need for Holistic Human-Centered Conceptualizations

Digital transformation promotes methods and disciplines that provide insights into either human processes (e.g., psychology) or human-centered approaches (e.g., human-computer interaction, HCI). The keywords explainable AI (e.g., Goebel et al., 2018; Samek et al., 2019) or human-centered AI (e.g., Riedl, 2019; Xu, 2019) refer to ongoing efforts to increase explainability and intuitive usability in the context of both the development and the implementation of AI systems (e.g., Wittpahl, 2019; Kraus et al., 2021). Approaches so far (at least implicitly) tend to follow the idea of the *Homo Economicus* and are therefore often limited to the analysis of pragmatic qualities (usability), while hedonic, eudaimonic, and social aspects—as well as professional and work-related aspects—are neglected (Carolus & Wienrich, 2021; Wienrich et al., 2022). Thus, this conception of the human limits the understanding of human-technology interactions as it neglects psychological and HCI knowledge. Only when this knowledge is taken into account do we adequately reflect the effects of human experiences, actions, and feelings. Consequently, skills and metacognitive competencies can be regarded as equally relevant as feelings of meaning and self-actualization (e.g., Mekler & Hornbæk, 2016).

AI-based technologies change the conceptualization of technology in general. The increasingly intelligent and interactive functions of AI turn what were once digital tools into (social) counterparts and interaction partners (Carolus et al., 2019; for a recent overview: Li & Suh, 2021; see also Reeves & Nass, 1996; Carolus et al., 2021; Wienrich et al., 2021).

The wider concept of user experience, which is enriched with emotions, perceptions, preferences, physiological and psychological reactions, behaviors, and performances occurring before, during, and after the interaction with technology, is regarded as essentially important for the consideration of usability in the professional work environment (Bargas-Avila & Hornbæk, 2011; Hassenzahl et al., 2010; Pataki et al., 2006). However, the current state of research reveals desiderata in terms of both the more comprehensive conceptualization of user experience and a widely accepted set of holistic and human-centered criteria for the evaluation of AI-based systems in the professional work environment.

2.3 Desideratum 3: The Need for Self-Determination

Comprehensive technological developments increase the complexity of systematic and valid investigations of human-technology interactions. Wienrich and Latoschik (2021) introduce the

eXtended AI approach as a new research heuristic that uses extended realities (XR; e.g., virtual reality, augmented reality) to enable systematic and valid investigations of AI systems and their effects. The basic idea is to use XR technology to create innovative testbeds reflecting complex AI interfaces and human-AI interactions. Utilizing rapid prototyping methods, *eXtended AI* reduces complexity and provides immersive experiences into certain fields of application. Moreover, various design spaces, easy accessibility, versatility, and tangible training possibilities are promising benefits of the *eXTended AI* approach. Possible forms of human-robot interaction (for instance, in logistics workplaces) were simulated virtually in a pilot study. Results show that different design features of the virtual robot contributed differently to fulfilling the diverse human needs at work. Significant gender differences were also identified. So far, the eXtended AI approach has only been studied in laboratory experiments with student samples—but not in professional work scenarios. As mentioned, there is a particular need for action in these professional contexts: While technologies and user requirements change rapidly, (future) users are rarely involved in change processes at an early stage. Approaches that allow low-threshold access to future technologies are still rare. This exclusion of users is highly problematic, as recent occupational psychology studies emphasize that limited opportunities to participate in the planning process lead to higher levels of dissatisfaction and increased workload (Carls et al., 2021). Conversely, the early involvement of users in the digital transformation process will positively affect the employees' work satisfaction.

To summarize, this study identifies a need for new participatory and human-centered approaches to introduce and implement AI in the workplace. These new approaches can contribute to the participatory, human-centered derivation of implementation requirements.

3 Concepts of Grasping AI Testbeds for Empowering Workers to Use and Competently Use AI

3.1 Addressing Desideratum 1: Design VALID and Straightforward Measures of AI Literacy

On the one hand, existing measures of AI literacy were mostly developed in the laboratory or under controlled conditions and with student samples who received course credit for participation. This resulted in large quantities of items and scales in a rather academic language. On the other hand, single-item measures and further less valid and reliable assessments are often used in studies in the field. Addressing both shortcomings, this study proposes a new strategy for developing practical measurements. In a top-down process, items and scales from existing scientific models are derived to create a scientifically reliable item pool. In a bottom-up process, the perspectives of domain experts and lay people who are affected by the introduction of AI systems at work are mapped to the item

pool. The resulting items are clustered into small modules and then adapted and reduced. The early involvement of people who are affected by technological change processes meets one of the essential requirements of various approaches (e.g., user-centered design, contextual design, value-sensitive design). Furthermore, the development of the measure of AI literacy is based on fundamental and applied research. The overall goal: short, comprehensive, valid, and reliable measures for AIL that fulfill both scientific (quality criteria) and practical requirements (increase the individuals' willingness to participate).

3.2 Addressing Desideratum 2: A Holistic Perspective on Human Beings at Work

As mentioned above, workspaces are highly associated with various human needs and motives. Consequently, the psychological perspective of this project emphasizes the importance of emotional and motivational aspects at work. For example, efficiency and effectiveness are equally important for the feeling of acceptance and community. Following this perspective, the project follows a holistic concept of the human being at work (going beyond the idea of the *Homo oeconomicus*). It considers emotional-motivational aspects as well as hedonic, eudaimonic, and social motives and needs. Moreover, the project will analyze the attribution of human characteristics to technology, which shape the users' expectations and interaction behaviors. Finally, the project considers the ongoing shift in perspectives on technology—from a perspective that considered technology as a tool to one that focuses on intelligent, interactive digital entities. Taken together, these three aspects will determine the criteria for requirements analyses, the design and development processes, and the practical implementations.

3.3 Addressing Desideratum 3: eXtended AI at Work

In everyday business, most of the employees who use technology are not involved in its development and implementation. However, excluding them at the early stage is risky. Mismatches between the systems and the employees' needs, which are detected after the system has been incorporated into the organizational processes, are cost intensive. Therefore, on the one hand, employee participation is promising. On the other hand, their perspective is limited in terms of their power of imagination and the validity of their projections. To close this gap, the present project proposes "graspable interactions" using the vast potential of XR. XR-based applications allow cost-efficient prototyping, which may vary in the degree of complexity and realism. Thus, XR is regarded as a powerful design space that can embody AI-based interaction partners without being bound to the limitations of conventional prototypes (e.g., physical environment, engineering challenges, costs). Furthermore, XR allows researchers to study heterogeneous user groups by adapting the prototypes to the individual's

degree of expertise or to different domains and tasks. Finally, XR provides a safe testbed without actual consequences in the real world (e.g., damages).

In sum, the project considers XR as a promising way to involve employees by providing AI experiences and explaining and illustrating the functionality and the consequences of the technology during the actual interaction. They can learn about AI, how to use and how to adapt the systems to their needs. Companies can trigger curiosity instead of fear and distrust by promoting participation and human-centeredness, development, and implementation.

4 Practical Examples Across Different Workplaces

4.1 Example 1: Robots in Logistics

Industrial robots are already used in various ways; they mostly work autonomously and at a safe distance from humans. In the future, robots are likely to collaborate more directly with humans (e.g., handing over work pieces). Hence, the design and behavior of robots (e.g., indicating social behavior or social intelligence) will likely have an impact on the acceptance and the human-robot work performance. XR testbeds can simulate the workplace and the robot in multiple versions. Employees can test and evaluate the different prototypes to contribute to further development—not only in line with economic and technical requirements but also in line with their human needs.

4.2 Example 2: Recommender Systems in Administration

AI can support administrative tasks in many ways. Against the background of the current legal situation, however, the human being is ultimately responsible for the decision. Therefore, the question arises of how AI can support human employees by providing information that is both relevant and accepted. Since human decisions are subject to numerous psychological biases, it is necessary to consider, for example, which interface and which information presentation lead to the best results, hence meeting the requirements of both the best decision and the human needs. In XR, different recommender systems can be tested and evaluated to contribute to further development. In addition to the administration, other domains are conceivable (e.g., medicine, law, insurance).

4.3 Example 3: Speech-Based Systems in Customer Service

Chatbots and voice assistants provide increasing support in customer service, mostly in addition to human consultants. Research shows that the design of the entities has a significant impact on the perception, acceptance, and trust of the systems and the service. In XR, different systems with

different designs and outward appearances can be tested and evaluated to contribute to further development.

5 Conclusion and Contribution

This paper emphasizes the urgency of transferring theoretical knowledge about critical demands for the development and implementation of human-centered AI systems into professional practice. While current public and political discussions tend to be limited to the business processes of only a few high-tech global players, this project focuses on people who work or will work with digital entities and AI systems. Moreover, the project incorporates a psychological perspective on increasingly digitalized workplaces, which are defined as essentially social spaces associated with multiple basic human needs. Thus, the employees' perspective—going beyond effectiveness and efficiency—becomes more complex and more important. Digital change processes of workplaces affect human beings, who are driven by their human needs. Additionally, employees come from different backgrounds and are equipped with different prior knowledge in computer science, resulting in interindividual differences in terms of their expectations from AI (dystopian fears vs. overestimating benefits), their participation and involvement at work, and their feelings of digital inequalities and disparities.

Summarizing the status quo, the project focuses on three desiderata: (1) the lack of valid, reliable, and practical measures of AIL, (2) untapped potential that arises from a holistic view of human-technology interaction that integrates human information processing and need structures, and (3) benefits of participative processes in development and implementation. These desiderata strongly contradict the demands of national and international AI strategies and endanger the safety of people and technology, job satisfaction and commitment, and value creation. To counteract this, our project proposes new strategies for socio-technical education and pedagogy in the context of work. It translates the critical demands and theoretical considerations from both science and politics into professional practice following the demands of the German federal government to contribute to the various layers of a socially responsible digital transformation.

6 Acknowledgments

This research has been funded by the German Federal Ministry of Labor and Social Affairs in the project AIL AT WORK (project number DKI.00.00030.21).

7 REFERENCES

1. Bargas-Avila, J. A., & Hornbæk, K. (2011). Old wine in new bottles or novel challenges: A critical analysis of empirical studies of user experience [Conference paper]. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/1978942.1979336
2. Bundesministerium für Bildung und Forschung. (n.d.). *Nationale Strategie für Künstliche Intelligenz* [National strategy for artificial intelligence]. https://www.ki-strategie-deutschland.de/home.html
3. Carls, K., Gehrken, H., Kuhlmann, M., Thamm, L., & Splett, B. (2021). Digitalisierung, Arbeit und Gesundheit [Digitalisation, work and health]. In K-P. Buss, M. Kuhlmann, M. Weißmann, H. Wolf & B. Apitzsch (Eds.), *Digitalisierung und Arbeit – Triebkräfte, Arbeitsfolgen, Regulierung* (pp. 235-273). Campus.
4. Carolus, A., Binder, J. F., Muench, R., Schmidt, C., Schneider, F., & Buglass, S. L. (2019). Smartphones as digital companions: Characterizing the relationship between users and their phones. *New Media & Society, 21*(4), 914-938. https://doi.org/10.1177/1461444818817074
5. Carolus, A., & Wienrich, C. (2021). Towards a holistic approach and measurement of humans interacting with speech-based technology. In C. Carolus, C. Wienrich & I. Siebert (Eds.), *1st AI-debate workshop: Workshop establishing an interdisciplinary perspective on speech-based technology* (pp. 39-41). Otto von Guericke University Magdeburg. http://dx.doi.org/10.25673/38471
6. Carolus, A., Wienrich, C., Törke, A., Friedel, T., Schwietering, C., & Sperzel, M. (2021). 'Alexa, I feel for you!'-Observers' empathetic reactions towards a conversational agent. *Frontiers in Computer Science, 3*, Article e682982. https://doi.org/10.3389/fcomp.2021.682982
7. European Commission. (2017). *Special Eurobarometer 460: Attitudes towards the impact of digitisation and automation on daily life*. https://ec.europa.eu/jrc/communities/sites/jrccties/files/ebs_460_en.pdf
8. Goebel, R., Chander, A., Holzinger, K., Lecue, F., Akata, Z., Stumpf, S., Kieseberg, P., & Holzinger, A. (2018). Explainable AI: The new 42? In A. Holzinger, P. Kieseberg, A. Min Tjoa & E. Weippl (Eds.), *Lecture notes in computer science: Vol. 11015. Machine learning and knowledge extraction* (pp. 295-303). Springer. https://doi.org/10.1007/978-3-319-99740-7_21
9. Hadan, H., & Patil, S. (2020). Understanding perceptions of smart devices. In M. Bernhard, A. Bracciali, L. J. Camp, S. Matsuo, A. Maurushat, P. B. Rønne & M. Sala (Eds.), *Lecture notes in computer science: Vol. 12063. Financial cryptography and data security* (pp. 102-121). Springer. https://doi.org/10.1007/978-3-030-54455-3_8
10. Hassenzahl, M., Diefenbach, S., & Göritz, A. (2010). Needs, affect, and interactive products – Facets of user experience. *Interacting with Computers 22*(5), 353–362. https://doi.org/10.1016/j.intcom.2010.04.002
11. Jenkins, H., Clinton, K., Purushotma, R., Robison, A. J. & Weigel, M. (2006). *Confronting the challenges of participatory culture: Media education for the 21st century*. The MacArthur Foundation. https://files.eric.ed.gov/fulltext/ED536086.pdf
12. Kelley, P. G., Yang, Y., Heldreth, C., Moessner, C., Sedley, A., Kramm, A., Newman, D. T., & Woodruff A. (2019). Exciting, useful, worrying, futuristic: Public perception of artificial intelligence in 8 countries. In M. Fourcade, B. Kuipers, S. Lazar, & D. Mulligan (Eds.), *Proceedings of the 2021 AAAI/ACM conference on AI, ethics, and society* (pp. 627–637). Association for Computing Machinery. https://arxiv.org/abs/2001.00081

13. Kraus, O., Ganschow, L., Eisenträger, M., & Wischmann, S. (2021). *Erklärbare KI. Anforderungen, Anwendungsfälle und Lösungen* [Explainable AI requirements, use cases and solutions]. Bundesministeriums für Wirtschaft und Energie. https://www.digitale-technologien.de/DT/Redaktion/DE/Downloads/Publikation/KI-Inno/2021/Studie_Erklaerbare_KI.pdf;jsessionid=C94CE1AF540F11B6D64E3EB19A61A2AF?__blob=publicationFile&v=9

14. Lau, J., Zimmerman, B., & Schaub, F. (2018). Alexa, are you listening? Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proceedings of the ACM on Human-Computer Interaction, 2*(CSCW), Article 102. https://doi.org/10.1145/3274371

15. Law, N., Woo, D., de la Torre, J., & Wong, G. (2018). A global framework of reference on digital literacy skills for indicator 4.4.2 (Information paper No. 51; p. 146). UNESCO Institute for Statistics. http://uis.unesco.org/sites/default/files/documents/ip51-global-framework-reference-digital-literacy-skills-2018-en.pdf

16. Lexparency.org. (n.d.). *Art. 14 AI-regulation (proposal) - Human oversight.* Retrieved September 1, 2022, from https://lexparency.org/eu/52021PC0206/ART_14/

17. Li, M., & Suh, A. (2021). Machinelike or humanlike? A literature review of anthropomorphism in AI-enabled technology. In T. X. Bui (Ed.), *Proceedings of the 54th Hawaii International Conference on system sciences* (pp. 4053-4062). Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2021.493

18. Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations [Conference paper]. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3313831.3376727

19. Madiega, T. (2022). *Artificial intelligence act*. European Parliamentary Research Service. https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf

20. Mekler, E. D. & Hornbæk, K. (2016). Momentary pleasure or lasting meaning? Distinguishing eudaimonic and hedonic user experiences [Conference paper]. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. http://dx.doi.org/10.1145/2858036.2858225

21. Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2*, Article e100041. https://doi.org/10.1016/j.caeai.2021.100041

22. Ng, W. (2012). Can we teach digital natives digital literacy? *Computer & Education, 59*(3), 1065–1078. https://doi.org/10.1016/j.compedu.2012.04.016

23. Pataki, K., Sachse, K., Prümper, J., & Thüring, M. (2006). ISONORM 9241/10-S: Kurzfragebogen zur Software-Evaluation [ISONORM 9241/10-S: Short questionnaire for software evaluation]. In F. Lösel & D. Bender (Eds.), *Kongress der Deutschen Gesellschaft für Psychologie (45.) – Humane Zukunft gestalten* (pp. 258-259). Pabst Science Publishers.

24. Porat, E., Blau, I., & Barak, A. (2018). Measuring digital literacies: Junior high-school students' perceived competencies versus actual performance. *Computer & Education, 126*, 23–36. https://doi.org/10.1016/j.compedu.2018.06.030

25. Reeves, B. & Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people. Cambridge University Press

26. Riedl, M. O. (2019). Human-centered artificial intelligence and machine learning. *Human Behavior and Emerging Technologies, 1*(1), 33-36. https://doi.org/10.1002/hbe2.117

27. Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K. R. (Eds.). (2019). *Lecture notes in computer science: Vol. 11700. Explainable AI: interpreting, explaining and visualizing deep learning*. Springer Cham. https://doi.org/10.1007/978-3-030-28954-6

28. Wienrich, C., & Carolus, A. (2021). Development of an instrument to measure conceptualizations and competencies about conversational agents on the example of smart speakers. *Frontiers in Computer Science, 3*, Article e685277. https://doi.org/10.3389/fcomp.2021.685277

29. Wienrich, C., Carolus, A., Augustin, Y., & Markus, A. (2022). *AI Literacy: Kompetenzdimensionen und Einflussfaktoren im Kontext von Arbeit* [Working paper] [AI Literacy: Competence dimensions and influencing factors in the context of work]. KI-Observatorium des Bundesministeriums für Arbeit und Soziales. https://www.denkfabrik-bmas.de/fileadmin/Downloads/Publikationen/AI_Literacy_Kompetenzdimensionen_und_Einflussfaktoren_im_Kontext_von_Arbeit.pdf

30. Wienrich, C., & Latoschik, M. E. (2021). Extended artificial intelligence: New prospects of human-AI interaction research. *Frontiers in Virtual Reality, 2*, Article e686783. https://doi.org/10.3389/frvir.2021.686783

31. Wienrich, C., Reitelbach, C., & Carolus, A. (2021). The trustworthiness of voice assistants in the context of healthcare investigating the effect of perceived expertise on the trustworthiness of voice assistants, providers, data receivers, and automatic speech recognition. *Frontiers in Computer Science, 3*, Article e685250. https://doi.org/10.3389/fcomp.2021.685250

32. Wittpahl, V. (2019). *Künstliche Intelligenz: Technologien | Anwendung | Gesellschaft* [Artificial intelligence: Technologies | application | society]. Springer. https://doi.org/10.1007/978-3-662-58042-4

33. Xu, W. (2019). Toward human-centered AI: A perspective from human-computer interaction. *Interactions, 26*(4), 42-46. https://doi.org/10.1145/3328485

34. Zeng, E., Mare, S., & Roesner, F. (2017). End user security and privacy concerns with smart homes. In M. E. Zurko, S. Chiasson & M. Smith (Eds.), *Proceedings of the thirteenth symposium on usable privacy and security* (pp. 65-80). USENIX Association. https://www.usenix.org/system/files/conference/soups2017/soups2017-zeng.pdf

35. Zhang, B., & Dafoe, A. (2019). *Artificial intelligence: American attitudes and trends*. Center for the Governance of AI, Future of Humanity Institute, University of Oxford. http://dx.doi.org/10.2139/ssrn.3312874