JOURNAL OF DIGITAL SOCIAL RESEARCH

VOL 7 : NO 1 : 2025

JOURNAL OF DIGITAL SOCIAL RESEARCH

VOL 7 : NO 1 : 2025

JOURNAL METADATA

ISSN
2003–1998

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

WEB
www.jdsr.io

ABOUT JDSR

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

CONTACT US

E-MAIL
editor@jdsr.se

JDSR 7(1) (2025) 1-19

JOURNAL OF DIGITAL SOCIAL RESEARCH

10.33621/jdsr.v7i148805

Digital transformation and research infrastructures

Promises and challenges of data-driven research in a Swedish context

Stefan Gelfgren and Coppélie Cocq

Umeå University, Sweden

🖂 stefan.gelfgren@umu.se

Abstract

Society is transforming due to changes in demographics, the environment, and technology, and thus faces multiple challenges. In this context, data coordination and access, collectively referred to as the digital transformation, are key to addressing anticipated societal tensions.

This interview-based qualitative study focuses on how researchers responsible for large-scale population-based research infrastructure view the opportunities and dilemmas in play in the intersection between data and personal privacy. The objective is to look beyond the glossy formulations of official strategy documents to see how the digital transformation (more specifically, data-driven research) is perceived from the active researcher's point of view, and what the intellectual negotiation process is like. What is of interest here is how the accessibility of register data is legitimized, and what developments and significant changes are simultaneously taking place. The research questions are:

1) How does the research community acknowledge the tensions and dilemmas between the possible risks and harms of large-scale, data-driven, population-based research, and its potential benefits?

2) How are the accessibility and coordination of research data justified and discussed by the research community, given the risks and potential, in relation to political and societal goals and policies?

With the contemporary Swedish research context as a point of departure, these research questions are addressed based on policy documents about digitalization, and on interviews with researchers.

Keywords: Digital transformation; digital humanities; surveillance culture; data-driven research; research ethics.

1. Introduction

Today, the opportunities to use data to track and analyze processes and behaviors are enormous, growing, and promising in many ways, but they are not unproblematic. These opportunities are embraced at a large scale through different means, by different actors, and with different agendas.

This article discusses and analyzes how personal data in registers and databases are used and expected to be used in creating new knowledge. The focus is mainly on social sciences, humanities, and health research, and how it can help find solutions to pressing problems in politics, culture, and health. This article examines the tensions between the accessibility of data in relation to research, on one hand, and the legal and ethical limitations of this accessibility, as perceived and voiced by representatives of the research community, on the other.

The point of departure of this article is the contemporary Swedish research context. The article is based on the reading of policy documents addressing digitalization, using them to establish the Swedish framework, and on five semi-structured interviews with researchers involved, at a national leadership level, in large-scale population-based databases. It addresses the following research questions:

1) How does the research community acknowledge the tensions and dilemmas between the possible risks and harms of large-scale, data-driven, population-based research, and its potential benefits?

2) How are the accessibility and coordination of research data justified and discussed by the research community, given the risks and potential, in relation to political and societal goals and policies?

The objective of this study is to understand the context, prerequisites, and potential development of population-based, data-driven research. The potentials, risks, limitations, and legitimization of data-driven research are approached from the perspective of active research leaders in relation to policies and agendas advocating digitalization. In focus is how they, in their role as researchers, need to negotiate hopes, fears, legislation, curiosity, and research ethics.

With this article, we contribute a humanistic, qualitative, and phenomenological perspective to the discussion of how the actual users of the data, the researchers, relate to the digital transformation. This enables us to discuss the implications and impacts of this research development concerning issues such as privacy, personal integrity, and legal frameworks. This stands in contrast to research focusing on, for example, law, science and policy, or science and technology studies approaches, even though this article is informed by work done within these disciplines.

As implied by the formulation of the concept, data-driven research is usually driven by data rather than by research questions. Regarding the data and databases of interest here, the causality runs both ways: new and larger datasets enable researchers to ask new questions, while attempts to answer new questions (some of which are undefined beforehand) drive the compilation of new datasets. This is a development that will continue with ongoing developments within artificial intelligence (AI).

2. Point of departure: Digitalization and digital transformation

As noted by, for example, Sadowski (2019), data have become essential for contemporary society—for commercial enterprises as well as governments: "just as we expect corporations to be profit-driven, we should now expect organizations to be data-driven; that is, the drive to accumulate data now propels new ways of doing business and governance" (p. 1). The abundance of data in combination with the development of increasing computational power and AI enable new opportunities to conduct data-driven research. For example, Vey et al. (2017) claimed that "we are at the beginning of a revolution that is fundamentally changing the way we live and work, the so-called Fourth Industrial Revolution" (p. 23).

Similar arguments have also been made by, for example, the UN Development Programme (2019), which has claimed that "emerging [digital] technologies have the potential to advance sustainability and to lead to better development work” (p. 6). The OECD (2019) has similarly stated that it should “enhance access to data to drive digital innovation [and] promote interoperable privacy regimes to facilitate cross-border data flows” (p. 3). In its strategic research agenda, the European Commission (2019) has stated that issues related to future businesses, climate, education, health, security, etc., depend on the digital transformation and how it is handled.

The impact of these developments on research has been discussed in previous studies. In their *Humanities World Report*, for instance, Holm et al. (2015) identified massive and complex data as a major research area where digital research is addressed by humanities scholars (pp. 68–72). Pappalardo et al. (2021) observed that data-driven science is “changing the way research is performed” (p. 261), arguing that a new paradigm is emerging. The abundance of data is changing research in and across different disciplines, such as geography (Miller & Goodchild, 2015), medical chemistry (Lusher et al., 2014), and psychology (Jack et al., 2018). In recent years, we have seen major advances in AI, not least concerning the publication of the ChatGPT bot (and other generative AI services), spurring discussions of the opportunities and complications presented by contemporary technological developments (see, e.g., the petition from March 2023 to “Pause Giant AI Experiments: An Open Letter”, initially signed by a thousand representatives of areas such as the tech industry and academia; Future of Life Institute, 2023).

Our interpretations are also informed by mediatization scholars such as Couldry and Hepp (2017), who observed that it is mainly private companies pushing the digitization, and ultimately the datafication (drawing on van Dijck, 2014), of society. In line with this, we note how the companies pushing digitalization will simultaneously own the data we all create, and that this raises questions of, for example, personal integrity and data ownership. Scholars such as Lyon (2017, 2018) and Zuboff (2019) view this development along similar lines, claiming that we live in a “culture of surveillance” (Lyon, 2018) or in a system of “surveillance capitalism” (Zuboff 2019). These scholars have seen how the digital footprints we all leave behind flow between, and are used by, businesses, banking systems, welfare authorities, the police and military, etc. (cf. Sadowski, 2019). Data are thus intertwined with our individual lives and behaviors, affecting our integrity and our perception of ourselves and our society. In this regard, Couldry and Hepp (2017) have claimed that we now live in a time of deep mediatization, meaning that the world and our relations are understood through media and data owned by companies and authorities, not by individuals.

A world of data indeed also offers new opportunities for research, but it also entails some rather problematic issues regarding, for example, personal privacy and legitimization.

This article focuses on the tensions between official digitalization agendas and concrete research practices, between the hopes associated with the digitization of society (“the digital transformation”) and the related dilemmas of the datafication of society. More specifically, we examine how Swedish researchers responsible for large-scale population-based databases and related research infrastructure view the opportunities and dilemmas in play today, how research access to data is legitimized, and what developments and significant changes are occurring. The objective is to look beyond the glossy formulations of national and international strategy documents to see how the digital transformation (specifically regarding data-driven research) is perceived from the active researcher’s point of view.

3. The Swedish case: The researchers’ framework

Along with the other Nordic countries, Sweden has a long history of compiling data about its citizens (see, e.g., Andreassen et al., 2021; Ustek-Spilda & Alastalo, 2020, for an overview). For decades (even centuries), these data had mainly administrative and scientific purposes within the welfare state system, but today they have become an asset to exploit.

Similar discussions have been taking place in the Nordic countries. For instance, Tupasela et al. (2020) have observed, in the context of Finland and Denmark, the emergence of the notion and imaginary of a “Nordic data gold mine,” and that the “logics of accumulation … reconfigure how the sources of this [sic] data are considered and imagined" (p. 2); such a discursive construct can be seen as "a necessary precondition" for articulating data policies ((Åm et al., 2021, p. 291 in a Norwegian context; see also Frank, 2020). Reutter and Åm (2024) have studied Norwegian policy documents to see how the discourse of hopes and technological determinism has been formulated. They have observed the Norwegian government pushing the digital transformation, noting that "datafication is accompanied by widespread beliefs that collecting and analyzing data can generate information and knowledge necessary for optimizing daily practices or for improving decision-making" (p. 1-2) and that "governments act as facilitators of digital markets" (p. 2).

In Sweden as early as the 17th century, clergy of the Church of Sweden (a state church until January 2000) registered Swedish citizens and kept track of birth and death dates, confirmation and marriage dates, names of families and relatives, and the ability to read and to understand the fundamentals of Christian doctrine (which were interwoven with the social order). These registers have been digitized and are now the basis of unique and world-renowned databases. In 1947, Sweden was the first country to introduce personal identification numbers to cover its whole population, initially for taxation purposes. Now the number is used whenever someone interacts with various state authorities, healthcare institutions, insurance companies, and banks and when ordering/buying various services. This gives Sweden (as well as other Nordic countries, as mentioned above) a unique position for register-based research at an individual level, especially if it is coordinated with health, welfare, and taxation registers (to mention a few).

Another contextual aspect of significance is the high level of trust in Sweden, as in the other Nordic countries, compared with large parts of the world. In the Nordic context, Sønderskov and Dinesen (2016) found "strong evidence of institutional trust influencing social trust" based on their analysis of datasets from Denmark. The Nordic countries are said to be "remarkable with respect to high levels of both social trust and, to a lesser extent, institutional trust" (Sønderskov & Dinesen, 2016, p. 187; see also Delhey & Newton, 2005; Zmerli et al., 2007), which are conclusions drawn from data from the World Values Survey (WVS) and the European Values Survey (EVS). Among groups earning the highest levels of trust, we find researchers and academia. And researchers are concerned with maintaining the level of trust they receive from the citizens.

As early as the 1990s there was discussion of the high value of data based on the Swedish resident population, but ethical concerns were also raised early on concerning surveillance, and issues of privacy and integrity. "Integrity issues with the IT age were discussed by both politicians and journalists, and the [governmental] IT-Commission was no exception. But it never triumphed over the belief that problems would be overcome" (Bennesved, 2024). This line of reasoning still resonates in contemporary discussions, as we will see below.

4. Material and methods

This article is based on both Swedish policy documents, which introduce the overall framework and give context, and interviews with research leaders, which add depth and complexity to the discussion. These two levels are then contrasted in the discussion.

The first part of our material thus consists of strategy documents and agendas published in the last decade that address the process of digitalization at a national level. Here we have focused mainly on the Swedish government and the Swedish Research Council (Vetenskapsrådet), the main authority through which the government channels research funding.

Close reading was used as a first step in order to identify patterns and central themes in these policy documents, especially formulations related to the anticipated digital transformation. These policy documents provide the baseline to which the interviews then are contrasted.

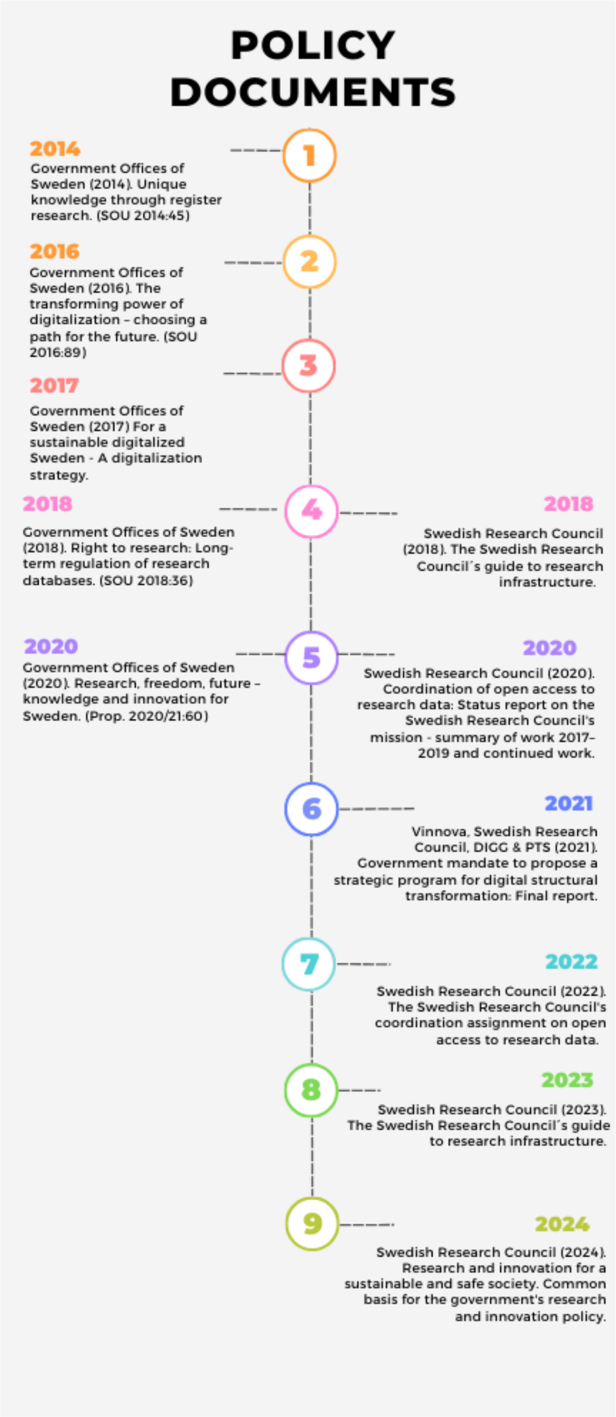

POLICY DOCUMENTS
1
2014
Government Offices of Sweden (2014). Unique knowledge through register research. (SOU 2014:45)
2
2016
Government Offices of Sweden (2016). The transforming power of digitalization – choosing a path for the future. (SOU 2016:89)
3
2017
Government Offices of Sweden (2017) For a sustainable digitalized Sweden - A digitalization strategy.
4
2018
Government Offices of Sweden (2018). Right to research: Long-term regulation of research databases. (SOU 2018:36)
2018
Swedish Research Council (2018). The Swedish Research Council´s guide to research infrastructure.
5
2020
Government Offices of Sweden (2020). Research, freedom, future – knowledge and innovation for Sweden. (Prop. 2020/21:60)
2020
Swedish Research Council (2020). Coordination of open access to research data: Status report on the Swedish Research Council's mission - summary of work 2017-2019 and continued work.
6
2021
Vinnova, Swedish Research Council, DIGG & PTS (2021). Government mandate to propose a strategic program for digital structural transformation: Final report.
7
2022
Swedish Research Council (2022). The Swedish Research Council's coordination assignment on open access to research data.
8
2023
Swedish Research Council (2023). The Swedish Research Council´s guide to research infrastructure.
9
2024
Swedish Research Council (2024). Research and innovation for a sustainable and safe society. Common basis for the government's research and innovation policy.

Figure 1. Overview of the documents analyzed (Authors' illustration).

To study tensions and dilemmas in relation to opportunities and challenges within the research community, we conducted in-depth interviews with established and active senior scholars involved in policy and decision-making processes at the national level. These interviews took the Swedish case as the point of departure. The scholars interviewed for this article were either heads of databases and related research infrastructure and/or were involved in policy and strategy work—all in disciplines that handle data from individuals, specifically the social sciences, humanities, and health research, and at the intersection between these traditional disciplines. They were also active researchers using the data obtained, for example, through and from the research infrastructure they led. They had extensive

https://doi.org/ 10.33621/jdsr.v7i148805

experience (30+ years) in their fields, and were chosen to cover a large interdisciplinary field. Gender and geographical distribution were also taken into consideration.[1]

Five interviews were conducted during the spring of 2020, after informed consent was received. Three interviews were conducted through video and two were conducted face-to-face; in all five cases, only audio-recordings were made (as preferred by the interviewees) and stored on the interviewer's computer. The interviews were conducted and transcribed in Swedish, and the translations into English are our own. As active researchers themselves, all participants were used to interviews, to the video format, and to being recorded, so no specific methodological problems were encountered. The interviews lasted 40–60 minutes and centered on questions related to the topic of this study, i.e., the opportunities and dilemmas associated with population-based databases and registry-based research. Questions were asked about issues such as the potential for doing research with large-scale datasets, and the risk of using data in a possibly individually intrusive way. These questions were based on the analysis of policy documents articulating the potential and opportunities offered by the digital transformation. Opportunities, in this context, included hopes and risks. Dilemmas emerged in the interviews when discussing the application and implementation (as a fact or effort) of the digital transformation in research.

Questions were asked about what opportunities there were and how they could be better utilized, about potential ethical dilemmas, the future, and perceived limitations. All interviewees expressed the desire that more researchers would use the register-based data, but all interviewees acknowledged that there were risks connected to the increased use of such data. Thus, the tensions and dilemmas this article seeks to explore were prevalent in the interview material.

The interviews were transcribed and coded based on a thematic analysis (Gray, 2002; Riessman, 2007) aligned with our research questions and our theoretical understanding of the subject. Recurring themes in all interviews were identified as potentials, risks, legitimization, and limitations. These themes are further analyzed and discussed in the empirical section below.

5. Theoretical framework

For this study, we drew on thematic discourse analysis. Our focus was on reflections and ideas rather than on technology and research activities per se. The aim was, as mentioned, to study how the use of various forms of register data is negotiated in the field of data-driven research (in the social sciences, health studies, and the humanities). The negotiation process is situated within the all-embracing political and societal discourse, which pushes research toward a digital transformation that is expressed and materialized in the political, policy, and strategy agendas. To understand this framework, within which the research discourse is situated, we were inspired by science and technology studies (STS). According to the STS scholars Biljer et al. (2012), the discourse has gained "momentum" and is supported by "organizations and people committed by various interests to the system" (p. 70). We accordingly lean more toward the sociology of studies than toward science policy studies (cf. Gläser & Laudel, 2016).

However, given the positions of the interviewees, they were also part of formulating research agendas. Here we were inspired by, for example, STS scholar Geels (2002), who saw how technological development was formulated at a macro level (here, political discourse), negotiated at a meso level (here, research discourse), and implemented/applied at a micro level (omitted here because this study was not interested in the actual implementation process), and noted that all three levels were interrelated. We saw how the researchers formulated a research discourse within and related to the political discourse.

Our approach also acknowledges the significance of the social and cultural contexts in which technologies take shape and are implemented. Such an approach is illustrated, for instance, by boyd & Crawford (2012), who defined "[b]ig data as a cultural, technological and scholarly phenomenon that

[1] Sweden is a small country, and Swedish academia is even smaller, so due to matters related to confidentiality, it is impossible to further specify the participants' background and still guarantee anonymity.

rests on the interplay of (1) Technology … , (2) Analysis … (3) [and] Mythology" (p. 663). In line with this perspective, our article takes the theoretical stand that it is important to consider the role of "the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy" (boyd & Crawford, 2012, p. 663). The concept of imaginaries (Lyon, 2018) inspires this approach by emphasizing the influence of what we think about what we do (see, e.g., Samuelsson et al., 2023)—in this case how imaginaries influence discourses and research practices.

6. Digitalization in Swedish policy documents

As early as 2011, the Swedish government formed a commission to formulate a digital agenda for Sweden, in order to foster the potentials brought by the digital transformation (Government Offices of Sweden, 2011). The commission's final report in 2016 was overtly positive toward digitalization, beginning with the commission chair's following claim:

> We live in exciting and interesting times! Digitalization is the most transformative societal process since industrialization. This development allows us to do entirely new things and to perform activities we previously engaged in, in entirely new ways. Our knowledge and understanding of humanity, society, and the environment will be transformed by the opportunities provided by the analysis of large amounts of data. Industrialization led to the development of the welfare society we have today, providing more people with the opportunity for a good life. Digitalization has the potential to develop a democratic and sustainable welfare society that we can hardly imagine today. (Government Offices of Sweden, 2016)

Still, the report acknowledges claims that societal benefits must be balanced against issues related to privacy risks. The digitalization strategy issued by the Swedish government the year after the commission's 2017 report states that Sweden should aim to be the best in the world when it comes to realizing the potential of digitalization. There are major potential benefits for the economy, employment market, and democracy; there are risks, though, connected to trust, democracy, and personal integrity (Government Offices of Sweden, 2017).

This strategy, focusing on the potential, is also evident in research strategies. In Government Offices of Sweden reports from 2014 and 2018, opportunities for register-based research are surveyed to find legal space in which to conduct such research—to balance potentials and questions related to privacy (Government Offices of Sweden, 2014, 2018). For example, the 2018 report mentions that "Sweden has a world-leading position when it comes to statistics regarding living conditions and health. Due to our individual-based registers, there are rich opportunities to perform successful research based on these. … To fully exploit these unique prerequisites, the more efficient use of existing registers and databases is needed" (Government Offices of Sweden, 2018).

In a Swedish government investigation from 2021, the authors, four Swedish research authorities (Vinnova, Swedish Research Council, DIGG – Myndigheten för digital förvaltning/Agency for Digital Government, & PTS – Post- och telestyrelsen/Swedish Post and Telecom Authority), acknowledged the urgent need for Sweden to strategically coordinate and use the potential of the vast amount of data it has accumulated (Vinnova et al., 2021). The driving forces are a competitive and sustainable economy, high-quality research supporting innovation, a desire to address societal challenges, and a drive to achieve accessible public administration supporting innovation and participation (p. 5). The process of digitalization is supposed to be "green, competitive, and centered on humanity" (p. 12). A few possible negative aspects were noted, with data management and storage being mentioned as a challenge concerning issues of personal privacy: "Digitalization brings new challenges … related to managing and storing data in relation to individual integrity. These challenges need to be addressed for the individual to trust in a digital society" (p. 14).

Related to the unique Swedish registers (often at the individual level), the digitization and coordination of individuals, and the possibility of using the Swedish registers is mutual *trust* among the state, the authorities, and fellow citizens. This is a challenge, and the strategic program for the digital transformation states that "these challenges have to be addressed for the individual to feel trust in relation to digital society" (Vinnova et al., 2021, p. 13).

The governmental research funding bodies the Swedish Energy Agency, FORMAS, Forte, the Swedish National Space Agency, the Swedish Research Council, and Vinnova were commissioned by the government (Government Offices of Sweden, 2023) to formulate a basis for the government's future research and innovation policy. In their report, data-driven research was deemed important, but these funders simultaneously acknowledged related privacy issues.

Today the Swedish research agenda acknowledges that the current digital transformation is changing the foundations of society. In the *Guide to research infrastructure 2023* (Swedish Research Council, 2023), digitalization is emphasized as a key component of future research addressing societal challenges, referring to the governmental research agenda. The *Guide* states that this must be done while taking account of ethics and privacy concerns. This tension comes through in formulations such as the following:

> The increased need for longitudinal studies of individuals also raises the potential conflict between the requirement to protect the integrity of sensitive personal data and the need for and opportunities of open science, where data may also be made accessible during peer review of scientific publication. (Swedish Research Council, 2023, p. 51)

The *Guide to research infrastructure 2018* (Swedish Research Council, 2018) mentions that it is essential to ensure personal privacy in order to maintain trust in and the legitimacy of research. As illustrated in the above quotation, the guiding research agendas articulate a push toward more openness and coordination of research data through open-access policies and data management according to the findable, accessible, interoperable, and reusable (FAIR) principles (see, e.g., Swedish Research Council, 2020, 2023), following a directive to the Research Council from the government in 2017. A Swedish Research Council (2022) report on the accessibility and coordination of open data also mentions that issues of personal integrity must be taken into account.

The policy documents illustrate a clear push toward digitalization within the political discourse in Sweden, with the benefits being highlighted while more problematic aspects play a complementary but subordinate role. The main discourse articulated in the political agendas is that the process of digitalization will give Sweden advantages in the areas of business, welfare, the environment, etc., all expressed and materialized in Swedish business and research agendas, and that, if not utilized, Sweden's potential as a digitally transformed society will be lost. This view (or imaginary) of digitalization is articulated in mainly positive terms and in a future-oriented discourse that depicts a better society using a somewhat utopian rhetoric (cf. boyd & Crawford, 2012). This is the political framework the research community works within.

7. Negotiations within the research discourse

As shown above, the policy documents largely favor an increasing degree of digitalization to address contemporary challenges through realizing the potential of data-driven science—manifested in research agendas and the implementation of the FAIR principles (meaning that data should be Findable, Accessible, Interoperable, and Reusable). Here we want to contrast and problematize these generalizations, setting them against the voices of researchers as they come through in the interviews, taking the discussion from a meta to a meso level.

Four main recurring themes were identified in our thematic discourse analysis of the interviews. These themes are illustrative of, and highlight, the tensions in play within the research discourse and structure the presentation of our results, namely: “Legislations”, “Limitations”, “Legitimizations”, and “Potentials and Risks”.

7.1 Legislations and guidelines – “If you look strictly at the law…”

The frameworks referred to by the interviewees include mainly ethical vetting and GDPR legislation. These frameworks are mentioned both as necessary in order to regulate data use and how research is conducted, and as limiting and maladapted in relation to recent changes in contemporary data and their uses.

The Swedish Ethical Review Authority is the national agency that reviews and grants ethics permits. Researchers are required to apply for ethical vetting in certain cases, for instance, if the research project involves sensitive information (i.e., in the context of the Review Act, information concerning ethnicity, political opinions, religious or philosophical beliefs, union membership, health, sexual life or orientation, and genetic and biometric data). In addition, some specific laws and ordinances regulate access to register data. For instance, confidentiality legislation regulates data that are worthy of special protection and must be kept confidential. The main rule is to guarantee public access to information, and the legislation delimits under what conditions it may be possible to access data despite their confidentiality.

The legal framework that regulates data use today is, according to the interviewees, necessary for the control of data and registers, as mentioned, for instance, by Interviewee 5:

> That is the reason why we have ethical vetting – to see if the benefits of the research conducted are great enough to disclose the data. It is such a balance we need to strike.

However, some interviewees had observed that current legislation was not adapted to contemporary conditions. Several interviewees mentioned the Swedish Ethical Review Authority as an important instance for the research process, but also identified areas that need to be developed and adapted. Knowledge of and conditions for managing social media data constitute a weakness of the authority mentioned by Interviewees 1 and 5. Another interviewee, number 4, had experienced that expertise was lacking in disciplines outside health and medicine. We were also told that opportunities for linking/merging data were limited due to legislation:

> Now we hope for new legislation. We do not really know what’s going on. There have been a number of investigations of the matter. … Right now we do not have any good legislation in this area in Sweden. For example, there is no good legislation for research databases that allows you to build a research database and make it available as infrastructure. You do that anyway, but there are, if you look strictly at the law, no such opportunities. (Interviewee 2)

The same interviewee commented that new legislation was needed “above all, to adapt to a reality that actually already exists.”

Worth noting is that several interviewees comment how the legislation and regulatory frameworks regarding data have changed over the years. The interviewed researchers described how research and research infrastructure have been – and still are – a driving force behind the development of laws and codes of conduct for data management. Therefore, the current situation is not fundamentally new, even though the scale of possibilities for conducting data-driven research has immensely increased, which continues to (again) push the boundaries of regulations.

Questions concerning register data and access to research data have, in the past as well, emphasized the tensions between legislation and research (Interviewees 1, 2, 3). For instance, Interviewee 3 stated: “From having been a bit controversial and suspect, collecting health data has become quite uncontroversial.” This quotation illustrates not only a change, but also how our attitudes toward the acceptance of data collection and data use have shifted.

The main dilemma identified by the interviewees was that an ethical and legal framework is important and necessary. The legal framework keeps research within ethical bounds and balances possibilities with what is reasonable and relevant to do. Simultaneously, and paradoxically, the framework is very difficult to implement, or the implementation limits the possibilities that current data-driven research can offer.

7.2 Limitations – "Great potential, if only the framework is adjusted"

The nature of data has indeed changed, and today data are digital and come with the possibility for different datasets to be coordinated and thus made interoperable. The amount of data that is produced and that can be collected is unprecedented, and the technical possibilities accessible to researchers for coordinating and analyzing data are rapidly developing. As Interviewee 2 straightforwardly said, "I think it is impossible to stop this". These new conditions not only put light on the tensions between legislation and data collection/use as illustrated above, but also actualize the ongoing challenges that contemporary research meets and addresses, given that the digital transformation within the political discourse has gained momentum, according to Bijker et al. (2012).

Our interviewees provided several examples of how researchers have striven to advance and adapt frames to contemporary conditions and needs. Interviewee 2 said that one vision or goal was to make as much data as possible accessible, but that current legislation was a hindrance. The role of researchers in identifying needs for changes and adapting the legal framework is nothing new. The relationship among researchers' expectations, the newly emerging opportunities, and the limits on what can be done was mentioned in several interviews. Interviewee 1, commenting on the establishment of data infrastructure (over 10 years ago) for which they were responsible, said:

> I think that the Ethics Review Board had a bit of a hard time seeing how that would fit in with legislation and practice. … it ended with us actually having to come to the board and answer questions.

Another interviewee shared a similar experience:

> When I started looking for infrastructure funds, those who sat in [the responsible board of a funding agency] did not understand. They did not seem to understand what it was, that it was even a question of infrastructure. (Interviewee 4)

Several interviewees returned to the fact that they, and the needs of the research community, are not understood properly by the law and by the vetting boards that restrict the use of registers and accessible data. The guidelines for the law and the vetting boards are based on the needs of medical research and were established during a time when data was relatively scarce, they say. The threat to privacy and integrity is overestimated, according to Interviewee 1, because it is never in the interest of researchers to single out individuals: the primary interest of research is large-scale patterns. The potential to use data in registers is huge, and such data could be used even better. Registers are built of anonymized data, but of course if different datasets are combined, a qualitatively new register is created, which needs to be assessed by vetting boards and must comply with laws and regulations.

The interviewees argued, for example, that as social media data exist and are used by companies, such data should also be used in research. Another reason to use new data, such as social media data, is the difficulties in collecting data using traditional methods, since people are now less likely to complete surveys and the like. The same difficulties in collecting data in traditional ways were mentioned by Interviewee 3, who said that people "do not want to give out their phone numbers, and they do not answer the phone, and they do not complete questionnaires, and so on." Here other forms of data can be helpful by pushing research toward new datasets offering new possibilities.

It is in the possibility of combining data from different registers and data sources that the real potentials can be found, and as Interviewee 4 said:

> One should be clear about that, that it is this connection, or interoperability, that makes data interesting, and also extremely sensitive in terms of personal privacy. There are so many aspects, different facets of a person's life, that can be put together into something complete. Yes, anyone can understand that this is potentially very sensitive.

Therefore, it is also important to have a clear legal framework and for researchers to understand that their activities are based on citizen trust, as claimed by Interviewee 4. Trust is thus highly valued by the researchers, and the risk of compromising it is carefully considered, especially in a country where trust in research and researchers is high, as mentioned in the introductory sections.

The interviewees shared the opinion that there were immense opportunities to perform new and innovative research in connection with digitalization. Merging datasets of various forms and making them searchable and interoperable would enable qualitatively and quantitatively new research, which would be interesting from the point of view of curious researchers. New questions could be asked and answered faster, on a larger scale and in real time. In particular, data that people share via, for example, social media and health apps, offer new potentials if they are combined with existing data on demographics, welfare, and societal infrastructure. Still, the risks of doing so were acknowledged.

7.3 Legitimization – "If private companies do it, why not us?"

The interviewees were well aware of the pitfalls and risks related to the use of population-based databases, but at the same time, all were aware that such data are already used in compiling and analyzing huge datasets within and related to businesses and companies. Global actors such as Google, Facebook, and Amazon were emphasized in this regard. It is well known that these actors gather and sell data to anyone interested based on the content that individuals provide them (cf. Lyon, 2017, 2018; Zuboff, 2019). This is both theoretically known and practically experienced, and the interviewees mentioned, for example, that advertisements based on their search history would appear shortly after searching for or buying particular items online. Rhetorically, they asked whether it would not be better and safer if researchers used similar tools and methods to conduct research for the collective good instead of for profit based on doubtful business models. Academic research is guided by laws, ethical guidelines, and best practices, which are considered to constitute a bulwark against the misuse of data. For example, "if we already have data regarding our movements [through mobile phones and health apps], why not use it to do some research too?" (Interviewee 1). "So, other actors, commercial actors [Google and Amazon were mentioned] surveil us in a very, yes, more intrusive way than researchers do" (Interviewee 3).

From the researchers' perspective, the other main reason for building and using large datasets is, as hinted above, the advantages individuals and society can gain from the related research results in areas such as health, the environment, or "epidemiology … and [overall] from a state financial perspective, on how to use our resources as efficiently as possible" (Interviewee 4). By connecting and coordinating different datasets, it is possible to gain insights into increasingly complex problems, which of course is both relevant and tempting. Contemporary society is facing large problems concerning health (accentuated by the Covid-19 pandemic) and the environment (accentuated and discussed in relation to what are considered contemporary extreme weather conditions), and here register-based research can help us understand causes and find solutions. Merging welfare registers capturing longitudinal living conditions with health registers and social media data could be a huge asset for researchers asking and answering new questions (Interviewees 1, 3, and 4).

However, as one interviewee mused, data-driven research is also about the sheer curiosity of researchers and the human impulse to explore whatever possibilities there are: "If we have Mount Everest, we need to climb it!" (Interviewee 2). If there are multiple datasets that could be enhanced and enriched by coordinating them to answer new questions, there are always people who, out of sheer curiosity, will want to try.

https://doi.org/ 10.33621/jdsr.v7i148805

Finally, it was noted that quantitatively large datasets are always anonymized and that there is no interest in the individual persons from which the data are derived. Therefore, it is not even possible or in any way relevant to identify the individual persons behind the data. It is argued that the risk of compromising the privacy of individuals is low to non-existent.

The bottom line is that the benefits of using these kinds of data in research are much greater than the risks, and because businesses can and do use such data, why not use them in a responsible way guided by ethical guidelines and best practices?

As mentioned in the background section on the Swedish context, Sweden has a long tradition of procuring register data. This historical background was mentioned by several interviewees as a motivation for continuity or even as "a tradition" (Interviewees 1 and 3) in data harvesting, collecting, etc. Interviewee 3 mentioned that "in Sweden, we have a tradition of digitizing both the health and welfare systems, since we were early adopters in this regard."

The trust and responsibility conferred by citizens (Interviewees 1 and 2) also legitimize the use of register data. One interviewee emphasized that their work with databases and registers was based on "the trust we have received from society, i.e., to use these data in a good way" (Interviewee 1). Sweden is a high-trust country, so it is not surprising that this aspect is an important motivation (and prerequisite) for the development of registers and databases. Trust in researchers was also mentioned in contrast to other actors involved in collecting (big) data. For instance, we were told by Interviewee 2 that research institutions are better than private or commercial actors in this regard:

> [In another country] they have [name of a database], which is one of the really large population databases, which is then linked together with healthcare data. And there it is owned by a private company. And it does not feel so good. … So, I think, I believe very much in the public. I also believe in the structured accessibility of what research needs.

On a similar note, another interviewee shared a reflection about how "it is not research that is the problem here, but it is … it is, so to speak, states and commercial actors" (Interviewee 5).

Overall, the aspects of tradition, continuity, and trust, referring to the Swedish case, are considered important for motivating and legitimizing population-based research. The interviewees discussed legitimization in relation to potentials and to what other actors in the field do – especially in relation to the business sphere. Here the interviewees referred to the need for researchers to maintain people's trust in them, otherwise their research would lose legitimacy. These are key aspects of how pure research differs from, for example, the suspect interests of commercial actors and states in surveilling their citizens. Research ethics, the idea of doing research for the greater good of society, and the need to maintain a trustworthy position also legitimize the access to, and use of, these kinds of data.

7.4 Potent data and potential risks – "Potentially very invasive with regard to privacy concerns"

Although the potentials of compiling and using registers and databases for research outweigh the associated risks and threats, the interviewees discussed possible negative consequences. Some risks can be foreseen, but these risks are more related to actors other than researchers who have other agendas, such as commercial actors (mentioned above) and state-governed intelligence agencies.

The other side of the huge possibilities and potential of register-based research can be seen in the policy documents. The better the coordination and combination of datasets, the better the potential to do ground-breaking research; at the same time, the more careful the consideration of privacy issues should be. In Swedish register data, it is possible to reach "a very fine and detailed level," so research based on these data could "potentially [be] very invasive with regard to privacy concerns" (Interviewee 4). Here we are talking about the potential and potent combination of health data from hospitals and recurring health surveys, data on demographics and socioeconomic conditions, social media content, and data from wearables.

Traditionally, data have been collected at recurring intervals, giving only snippets of information limited to capturing conditions as specific points in time. However, with a more constant flow of data from, for example, social media and health apps, in combination with, for example, AI, language translation, and facial recognition, data will reach a new level of continuity and granularity. This will entail both greater potential and greater risk – the latter perhaps less in a research context, but more in the hands of malevolent states, businesses, and political actors, as mentioned in the interviews. Concerns were also raised regarding the possible use of such datasets to shape public opinion and consumer patterns (which is already taking place and was a main concern, for example, in the Cambridge Analytica scandal in 2018 and in the US Special Counsel Investigation in 2017–2019).

Social media data and data shared through, for example, health apps and services, were singled out as both extremely potent and risky: "The data we share about how we move, what we say, what we think, and what we download, these data are authentic data on a whole different level" (Interviewee 1). On the other hand, it is difficult to understand why people generally freely distribute such data, as Interviewee 5 observed, fascinated by the volume of data people share apparently without concern about how they could be used:

> People share so much data about themselves, and so much sensitive data about themselves, at the same time as there are so many conflicts concerning data in relation to research … It does not make sense! (Interviewee 5)

One interviewee also mentioned that datasets not intended to be out in the open are made accessible under the provisions of what is referred to as open access. Publishers demand that researchers publish their data in order to make the research process transparent, but, for example, the so-called quality registers (or health data registers) based on medical records were never intended to be public. We have a situation in which the research discourse is promoting open access, whereas:

> The quality registers were built by healthcare actors, intended to ensure quality in the healthcare system, but now the demands to open these registers are increasing, which the medical professionals are a bit reluctant to do. (Interviewee 3)

The question is how to draw the line between research interests and personal privacy, as formulated by Interviewee 5, when the potential also entails risks. Again, larger datasets enable interesting research regarding complex and relevant questions, but at some point, the sheer amount of coordinated data might become problematic. Data collected or shared for other purposes than originally intended, for example, data from health service quality registers or people's social media data, are, if combined, a potential risk, although simultaneously offering huge potential.

8. Discussion

The themes of research "Legislations", "Limitations", "Legitimization", and "Potentials and Risks" are factors the interviewees mentioned that influence the research discourse and are thus negotiated. The research discourse takes these factors into consideration while negotiating with the political discourse regarding the terms and conditions of the research discourse (see Figure 2). This form of negotiation illustrates how digitalization as a socio-technical phenomenon is perceived, articulated, and implemented according to both utopian imaginaries – i.e., how research and society will benefit from it – and cautious imaginaries emphasizing privacy (see, e.g., Lehtiniemi & Ruckenstein, 2019; Lyon, 2018; Tupasela et al., 2020).

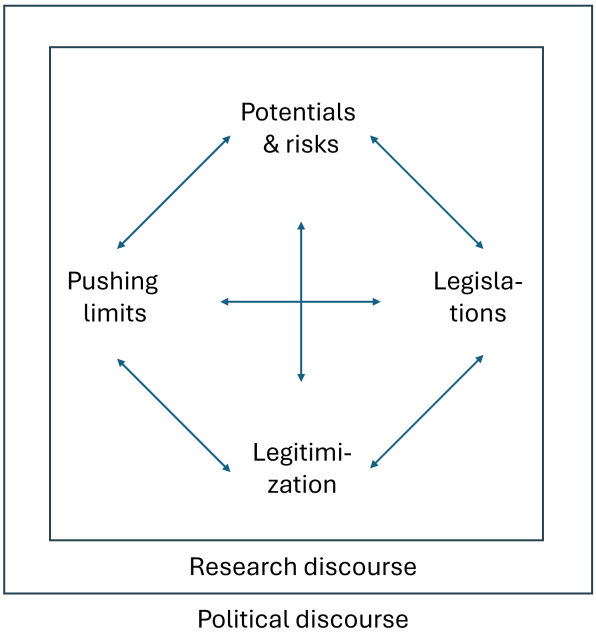


Figure 2. The negotiation process that occurs in the research discourse, related to the political discourse.

The political discourse, as mentioned in policy and strategy documents, encourages society and research to move toward increased digitalization. In Swedish policy documents, this is articulated as building on a unique tradition of collecting data about citizens—as an asset to explore and capitalize on – in line with the idea of a "Nordic data gold mine," as suggested by Tupasela et al. (2020).

However, taking account of discussions within the research community, as manifested in the interviews, it becomes clear that the digital transformation is not a predetermined one-way process that can be implemented without discussion of its consequences. Instead, it is a discourse that recalls the privacy paradox, which has been studied among ordinary citizens and users of digital media. Such a paradox, defined as the "discrepancy between individuals' intentions to protect their own privacy and how they behave in the marketplace" (Norberg et al., 2007, p. 101; see also, e.g., Kokolakis, 2017, for an overview), highlights the tensions arising when the user is "expected to trade the benefits that could be earned by data disclosure off against the costs that could arise from revealing his/her data" (Gerber et al., 2018, p. 229). More recent research indicates, however, that what is perceived as paradoxical at first sight can indeed be "partially interpreted and explained in terms of ethical and ideological considerations" (Cocq et al., 2020, p. 191).

Despite the overtly positive political discourse – both nationally and internationally – the interviewees expressed a more nuanced view of digitalization. The digital transformation of society in general, and of research in particular, was seen as bringing great potential and benefits, but also as needing to be balanced against anticipated risks (indeed, as noted in the policy documents but rarely dwelled on, or problematized, at length).

This discourse is present in all the above themes, and we find several contradictory aspects discussed in relation to the use of population-based research data of various forms. For example, the legal and ethical frameworks hinder researchers from doing whatever they want with the data. They are thus mentioned as important for ensuring that research questions are balanced against proper needs, relevance, and research ethics. The trust researchers feel they have from the citizens is also part of the discussion. Trust is valued and something they care about. Similarly, the interviewees saw huge potential in better computational power and new and interoperable data, but they also saw the legal framework as obsolete and inappropriate for the kind of research they wanted to do.

Ethical issues raised by data-driven technologies are being addressed and discussed in recent research, for instance, concerning the role and prerequisites of ethics committees reviewing research based on data-driven technologies in university contexts (Hine, 2021). Such research emphasizes the limitations and challenges that ethical committees encounter and underscores the need to develop effective systems of ethical governance. Also, Forgó et al. (2020) discussed the need for suitable infrastructural, organizational, and methodological principles when establishing ethical – legal frameworks in research. In a Swedish context, as illustrated and discussed under the themes of Legislations and Limitations, the work conducted by review boards and associated infrastructures in relation to ethics and legislation concerns understanding and adapting the current framework, which is being challenged by data-driven research.

The interviewees discussed the role of research as a curiosity-driven activity, and they were eager to find new solutions to existing and anticipated problems in research done for the greater good, and thus in line with both the research opportunities and potential breakthroughs that come with increased digitalization. Here the research discourse differs from, for example, the business discourse, which is primarily based on a profit-driven agenda. In the research discourse, the researchers discuss and negotiate risks associated with research conducted in relation to the very same process of digitalization.

The opportunities associated with data-driven research are balanced against its potential risks – i.e., addressing interesting and complex research questions versus threats to personal privacy. Ongoing and future research must be balanced and legitimized by invoking the great societal need for innovative research, and by comparing this need with what has already been done by other actors with commercial/financial interests. In contrast, the researcher's personal curiosity is a driving force that might push the boundaries of research.

9. Conclusions

Significant developments and changes have been occurring in relation to the use of research data, infrastructure, etc. This study casts light on the associated tensions and dilemmas, from the perspective of the research community, as these developments and changes have occurred in legislation and in practice. Recent generative AI developments highlight these tensions in relation to questions of the ownership of the data on which the AI models are trained (see, e.g., Lucchi, 2023; Samuelson, 2023).

Overall, we have a situation in which the political discourse has gained momentum and is pushing for digital transformation at the macro level. In the research community and research discourse, these tensions are articulated. Here we have aimed to nuance discussion of the digital transformation – a process that is not a predetermined one-way process, even though the political discourse has promoted the development of data-driven research, in line with the anticipated digital transformation of society. As the analysis of the policy documents clearly indicates, digital transformation and data-driven research are seen as keys to addressing contemporary challenges related to demographics, climate change, and democracy, also having the potential to give rise to new business models. The concept of open data in relation to technological advances is key to encountering the future, although issues such as personal privacy are acknowledged to be at stake.

Different interests and arguments were, however, discussed and balanced in relation to one another by the interviewed researchers. The potential to conduct new, important, and relevant research is acknowledged in the research discourse, but such research is not as straightforward or single-minded as is described in research policies and documents at the national and global levels. Analysis of the interviews shows that there are different reasons to legitimize the use of data and to coordinate and merge large datasets. The legitimization arguments expressed in our interviews centered on the facts that research using these forms of population-based data is beneficial for society and the greater good, that the risk of violating someone's privacy is low to none, and that it is better that such data be used by researchers working under ethical and legal guidelines than by businesses working for profit.

The forces in play concern the increased digitalization of society and what it means for research are, as we have shown, multiple. Policies, agendas, and political discourses are explicit, plain, and clear. The data considered here are also embraced by commercial actors, who are yet another driving force shaping how data are compiled and shared. However, we also see that the research community is a key actor in this process. In practice, the frameworks within which large databases and data infrastructure are developed and applied are constantly challenged by research and researchers, resulting in necessary adaptations.

This article has focused on the use of population-based data, discussing the contemporary and potential use of such data. However, to fully understand the opportunities and potentials of society's digital transformation – in research and in other areas of society – we must consider the owners and providers of such data, namely, the citizens. A related study conducted by our research group shows that many are concerned about their data being used without their consent, and thus adjust their online behavior to conceal their data or prevent it from being gathered (Cocq et al., 2020). The interviewees in this study also touched on the fact that it has become increasingly difficult to obtain data by having people complete questionnaires and voluntarily participate in research studies. Therefore, other data collection approaches are undertaken, and to obtain other forms of data, for example, through social media platforms. Our interviewees also mentioned that medical professionals who have compiled registers of health data (the quality registers) are hesitant to open these registers for research for other purposes than originally intended.

Ultimately, it is a matter of the legitimacy of the digital transformation, in research and elsewhere, and the question is whether potentially sensitive data should be allowed to "float around" and be used for various purposes other than originally intended as long as the intentions are good and relevant. This question was briefly touched on in the empirical material and selected policy documents, but to ensure continued trust and legitimacy in the future digital transformation of research, this article shows how these complex questions must be addressed more thoroughly.

Acknowledgements

This article is part of the project iAccept: Soft Surveillance – Between Acceptance and Resistance, funded by the Marcus and Amalia Wallenberg Foundation (MAW2016.0092).

References

Åm, H., Solbu, G., & Sørensen, K. H. (2021). The imagined scientist of science governance. *Social Studies of Science*, 51(2), 277-297. https://doi.org/10.1177/0306312720962573

Andreassen, R., Kaun, A., & Nikunen, K. (2021). Fostering the data welfare state: A Nordic perspective on datafication. *Nordicom Review*, 42(2), 207–223. https://doi.org/10.2478/nor-2021-0051

Bennesved, P. (2024). Surveillance, integrity and metadata in the information age The legacy of Sweden's ICT commission, 1994–2003. *Lychnos: Årsbok för idé- och lärdomshistoria,* 105–129. https://doi.org/10.48202/26116.

Bijker, W. E., Hughes, T. P., Pinch, T. J. (2012). *The social construction of technological systems: New directions in the sociology and history of technology* Cambridge, MA: MIT Press.

boyd, d., & Crawford, K. (2012). CRITICAL QUESTIONS FOR BIG DATA: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Cocq, C.; Gelfgren, S.; Samuelsson, L.; Enbom, J. (2020) Online Surveillance in a Swedish Context: Between acceptance and resistance. *Nordicom Review*, 41(2), 179–193. https://doi.org/10.2478/nor-2020-0022

Couldry, N. & Hepp, A. (2017). *The mediated construction of reality*. Cambridge: Polity.

Delhey, J. & Newton, K. (2005). Predicting cross-national levels of social trust: Global pattern or Nordic exceptionalism? *European Sociological Review*. 21(4). 311–327. https://doi.org/10.1093/esr/jci022

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

European Commission. (2019) *Orientations towards the first strategic plan for Horizon Europe*. Accessed 10 Feb. 2021 from https://research-and-innovation.ec.europa.eu/system/files/2019-12/ec_rtd_orientations-he-strategic-plan_122019.pdf

Forgó, N., Hänold, S., van den Hoven, J., Krügel, T., Lishchuk, I., Mahieu, R., Monreale, A., Pedreschi, D., Pratesi, F., & van Putten, D. (2020). An ethico-legal framework for social data science. *International Journal of Data Science and Analytics*, 11(4), 377–390. https://doi.org/10.1007/s41060-020-00211-7

Frank, F. (2020). When an Entire Country is a Cohort. *Science* 287, 2398-2399. https://www.science.org/doi/10.1126/science.287.5462.2398

Future of Life Institute. (2023, 22 mars). Pause Giant AI Experiments: An Open Letter, Accessed 2 Feb. 2024 from https://futureoflife.org/open-letter/pause-giant-ai-experiments/

Geels, F. W. (2002). Technological transitions as evolutionary reconfiguration processes: A multi-level perspective and a case-study. *Research Policy* 31(8). 1257–1274. https://doi.org/10.1016/S0048-7333(02)00062-8

Gerber, N., Gerber, P. & Volkamer, M. (2018), Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior. *Computers & Security*, (77). 226-261. https://doi.org/10.1016/j.cose.2018.04.002

Gläser J, Laudel G. Governing Science: How Science Policy Shapes Research Content. *European Journal of Sociology*. 2016;57(1):117-168. https://doi.org/10.1017/S0003975616000047

Government Offices of Sweden (2011). *IT i människans tjänst—en digital agenda för Sverige (dnr N2011/342/ITP)*. See https://www.regeringen.se/rattsliga-dokument/kommittedirektiv/2012/06/dir.-201261.

Government Offices of Sweden (2014). *Unik kunskap genom registerforskning [Unique knowledge through register research]*. (*SOU 2014:45)*. Accessed 3 Dec. 2021 from https://www.regeringen.se/contentassets/2e42baeaa76e4e918dd852f2a3e43fca/unik-kunskap-genom-registerforskning-sou-201445

Government Offices of Sweden. (2016). *Digitaliseringens transformerande kraft—vägval för framtiden [The transforming power of digitalization—choosing a path for the future] (SOU 2016:89)*. Accessed 15 Feb. from https://www.regeringen.se/contentassets/f7d07b214e2c459eb5757cea206e6701/sou-2016_89_webb.pdf

Government Offices of Sweden (2017) *För ett hållbart digitaliserat Sverige—En digitaliseringsstrategi [For a sustainable digitalized Sweden: A digitalization strategy]*. Accessed 8 Aug. 2023 from https://www.regeringen.se/49adea/contentassets/5429e024be6847fc907b786ab954228f/digitaliseringsstrategin_slutlig_170518-2.pdf

Government Offices of Sweden (2018). *Rätt att forska: Långsiktig reglering av forskningsdatabaser [Right to research: Long-term regulation of research databases]. (SOU 2018:36)*. Accessed 3 Dec. 2021 from https://www.regeringen.se/49c0cd/contentassets/2570b2abef8c48f084bbddb48ed2300b/sou-2018_36_webb.pdf

Government Offices of Sweden (2020) *Forskning, frihet, framtid—kunskap och innovation för Sverige [Research, freedom, future—knowledge and innovation for Sweden]*. Prop. 2020/21:60. Accessed 23 May 2023. https://www.regeringen.se/rattsliga-dokument/proposition/2020/12/forskning-frihet-framtid--kunskap-och-innovation-for-sverige/

Government Offices of Sweden (2021). *Regeringsuppdrag att föreslå ett strategiskt program för digital strukturomvandling*. Accessed 3 Dec. 2021 from https://www.regeringen.se/48e3a9/contentassets/d9408f8eee6342b2b4aba3c52072dfcb/uppdrag-att-foresla-utformning-av-ett-strategiskt-program-for-att-mota-och-leda-i-den-digitala-strukturomvandlingen.pdf

Government Offices of Sweden (2023). Uppdrag till forskningsfinansiärerna att inkomma med analyser som underlag till regeringens forskningspolitik. (Dnr U2023/01317) Accessed 15 Feb. 2024 from https://www.regeringen.se/contentassets/5eee3b9b4a32457ea9a5fae3cf2e0bbc/uppdrag-till-forskningsfinansiarerna-att-inkomma-med-analyser-till-regeringens-forsknings--och-innovationspolitik.pdf

Gray, A. (2002). *Research practice for cultural studies: Ethnographic methods and lived cultures*. London: Sage.

Hine, C. (2021). Evaluating the prospects for university-based ethical governance in artificial intelligence and data-driven innovation. *Research Ethics Review*, 17(4), 464–479. https://doi.org/10.1177/17470161211022790

Holm, P., Jarrick A., Scott D. (2015). *Humanities World Report*. Basingstoke: Palgrave Macmillan.

Jack, R. E., Crivelli, C., & Wheatley, T. (2018). Data-driven methods to diversify knowledge of human psychology. *Trends in cognitive sciences*, 22(1), 1–5. https://doi.org/10.1016/j.tics.2017.10.002

Kokolakis, S. (2017), Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, (64), 122–134. https://doi.org/10.1016/j.cose.2015.07.002

Lehtiniemi, T., & Ruckenstein, M. (2019). The social imaginaries of data activism. Big *Data & Society*, 6(1), Article 2053951718821146. https://doi.org/10.1177/2053951718821146

Lucchi, N. (2023). ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems. *European Journal of Risk Regulation*, 1–23. https://doi.org/10.1017/err.2023.59

Lusher, S. J., McGuire, R., van Schaik, R. C., Nicholson, C. D., & de Vlieg, J. (2014). Data-driven medicinal chemistry in the era of big data. *Drug Discovery Today*, *19*(7), 859–868. https://doi.org/10.1016/j.drudis.2013.12.004

Lyon, D. (2017). Surveillance culture: Engagement, exposure, and ethics in digital modernity. *International Journal of Communication*. 11, 824–842. https://ijoc.org/index.php/ijoc/article/view/5527

Lyon, D. (2018). *The culture of surveillance: Watching as a way of life*. Cambridge: Polity Press.

Miller, H. J., & Goodchild, M. F. (2015). Data-driven geography. *GeoJournal*, *80*(4), 449–461. https://doi.org/10.1007/s10708-014-9602-6

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs*, 41, 100–126. https://doi.org/10.1111/j.1745-6606.2006.00070.x

OECD. (2019). *Going Digital: Shaping Policies, Improving Lives*. Paris: OECD Publishing.

Pause Giant AI Experiments: An Open Letter. Accessed 4 April 2023 from https://futureoflife.org/open-letter/pause-giant-ai-experiments

Pappalardo, Grossi, V., & Pedreschi, D. (2021). Introduction to the special issue on social mining and big data ecosystem for open, responsible data science. *International journal of data science and analytics*, 11(4), 261–262. https://doi.org/10.1007/s41060-021-00253-5

Reutter, L. & Åm, H. (2024) Constructing the data economy: tracing expectations of value creation in policy documents, *Critical Policy Studies*. https://doi.org/10.1080/19460171.2023.2300436

Riessman, C.K. (2007). *Narrative Methods for the Human Sciences*. London: Sage.

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

Samuelson, P. (2023) Generative AI meets copyright. *Science* 381,158-161. https://www.science.org/doi/10.1126/science.adi0656

Samuelsson, L., Cocq, C., Gelfgren, S., Enbom, J. (2023). *Everyday Life in the Culture of Surveillance*. Nordicom, University of Gothenburg. https://doi.org/10.48335/9789188855732

Swedish Research Council (2018). *The Swedish Research Council´s guide to research infrastructure 2018*. Accessed 3 Dec. 2023 from https://www.vr.se/download/18.7f26360d16642e3af996ff/1555326273381/Vetenskapsraadets-guide-infrastrukturen-2018_VR_2018.pdf

Swedish Research Council (2020). *Samordning av öppen tillgång till forskningsdata: Statusrapport i Vetenskapsrådets uppdrag—summering av arbetet 2017–2019 och fortsatt arbete [Coordination of open access to research data: Status report on the Swedish Research Council's mission—summary of work 2017–2019 and continued work.]*. Accessed 3 Dec. 2023 from https://www.vr.se/download/18.47291b121711f04ce6e3bb/1585815456838/Samordning%20av%20oppen%20tillgang%20till%20forskningsdata_VR_2020.pdf

Swedish Research Council. (2022). *Vetenskapsrådets samordningsuppdrag om öppen tillgång till forskningsdata 2022 [The Swedish Research Council's coordination assignment on open access to research data].* Accessed 15 Feb. 2024 from https://www.vr.se/download/18.72c4495e17f44b64443b03a/1647009787100/Samordningsuppdrag%20om%20%C3%B6ppen%20tillg%C3%A5ng%20till%20forskningsdata%20VR%202022.pdf.

Swedish Research Council (2023). *The Swedish Research Council´s guide to research infrastructure 2023*. Accessed 15 Feb. 2024 from https://www.vr.se/download/18.14f2b85918c8ae70d9922c20/1704736135760/Guide%20to%20research%20infrastructure%202023.pdf

Swedish Research Council (2024). *Forskning och innovation för ett hållbart och säkert samhälle Gemensamt underlag till regeringens forsknings- och innovationspolitik från Energimyndigheten, Formas, Forte, Rymdstyrelsen, Vetenskapsrådet och Vinnova*. [Research and innovation for a sustainable and safe society. Common basis for the government's research and innovation policy] Accessed 15 feb 2024 from https://www.vr.se/download/18.1e171f1718bcc1830fea98/1699964201384/Forskning%20och%20innovation%20f%C3%B6r%20ett%20h%C3%A5llbart%20och%20s%C3%A4kert%20samh%C3%A4lle-2023.pdf

Sønderskov, K. M., & Dinesen. P. T. (2016). Trusting the state, trusting each other? The effect of institutional trust on social trust." *Political behavior* 38(1). 179–202. https://doi.org/10.1007/s11109-015-9322-8

Tupasela, A., Snell, K., & Tarkkala, H. (2020). The Nordic data imaginary. *Big Data & Society*, 7(1). https://doi.org/10.1177/2053951720907107

UNDP. (2019). *Future forward: UNDP digital strategy*. Accessed 10 Feb. 2021 from https://digitalstrategy.undp.org/documents/UNDP-digital-strategy-2019.pdf

Vey, K., Fandel-Meyer, T., Zipp, J. S., & Schneider, C. (2017). Learning & Development in Times of Digital Transformation: Facilitating a Culture of Change and Innovation. *International Journal of Advanced Corporate Learning (iJAC)*, *10*(1). 22–32. https://doi.org/10.3991/ijac.v10i1.6334

Vinnova, Vetenskapsrådet, DIGG & PTS. (2021). *Regeringsuppdrag att föreslå ett strategiskt program för digital strukturomvandling: Slutredovisning* [Government mandate to propose a strategic program for digital structural transformation: Final report] https://www.vinnova.se/contentassets/b6f628d9450642068ce283db0f16381d/rapport-ru-kraftsamling-for-digital-strukturomvandling.pdf

Zmerli, S., Newton, K., & Montero, J. R. (2007). Trust in people, confidence in political institutions, and satisfaction with democracy. In *Citizenship and involvement in European democracies: A comparative analysis*. Eds. J. W. van Deth, J. R. Montero, & A. Westholm. New York: Routledge.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. New York: Public Affairs.

https://doi.org/ 10.33621/jdsr.v7i148805

‘This is neither Swedish nor Western and doesn’t belong here’

Responses to retail stores’ social media advertisements addressing Ramadan

Johanna Sixtensson, Paula Mulinari and Carina Listerborn

Malmö University, Sweden

✉ johanna.sixtensson@mau.se

Abstract

This article analyses written online responses to Swedish retail stores’ social media advertisements broadly addressing the Muslim celebration of Ramadan. It is based on a selection of 19 social media advertisement posts that together generated a total of 2988 responses in discussion threads. The customer responsive comments are analysed through the theoretical lens of race and racism in the digital society and theories of everyday nationhood and nationalism. At large, the result shows that the social media platforms can be seen as facilitators of anti-Muslim racism. However, the advertisements and the responses to them, which express dislike of as well as support for the retailers, Muslim traditions and the Muslim community, illustrate a negotiation of nationhood which is characterized on the one side by racist anger and fear of loss of nation, and on the other side by support for inclusion. Inspired by the concept of ‘predatory inclusion’, the article argues that this paradoxical phenomenon illustrates both inclusion and exclusion. The retail stores’ social media platforms are not only spaces of hatred against Muslims but also a space in which resistance to anti-Muslim racism is articulated and where constructions of Swedishness are challenged.

Keywords: Social media platforms, Digital consumer spaces, Anti-Muslim racism and Islamophobia, Nation and nationhood, Exclusion and inclusion, Retailers’ advertisement

1. Introduction

> Today is the start of Ramadan, a precious time for Muslims around the globe. We want to celebrate this! The shop is stocked up and the prices are awesome. Welcome! 😍
>
> (Retail shop’s social media advertisement post)

> This is neither Swedish nor Western and doesn’t belong here.
>
> (Customer response)
>
> Open-minded and clever of you to think about other cultures. Thank you! ❤️
>
> (Customer response)

Be it of foods or other retail goods, consumption is a central and habitual everyday practice for most people. Today, many consumer practices take place online and the digital arena is a space where retailers communicate with potential customers (Rydström, 2024). As part of this development, Swedish retail store chains have recently begun acknowledging the Muslim celebration of Ramadan on their social media websites to attract customers to their stores. The advertisements are generally informal posts on Facebook, like the above excerpt; for instance wishing happy celebrations and emphasizing certain items or foods and offering discounts. In this article we are predominantly interested in examining the (sometimes vast number of) customer responses to such initiatives. Such responses are articulated in discussion threads below the retail store’s social media announcements, as exemplified in the second and third examples above.

That Muslims are racialised and exposed as customers in physical retail settings has been illustrated in Alkayyali’s (2019) study, which shows that Muslim women who wear headscarves experience racial profiling such as invisibilization and objectification as consumers in French retail settings. One coping strategy revealed in the material is that the women choose to do their shopping online to avoid such experiences. The Internet cannot be seen as a space free of racism, even if the subject of Internet (still) tends to be separated from the subject of race: ‘the mechanisms of color-blind racism are interwoven in fantasies of the Internet as a raceless utopia’ (Daniels, 2015, p. 1388, see also Matamoros-Fernández, 2017).

Employing Tressie McMillan Cottom’s (2020) sociology of race and racism in the digital society and her understanding of platform capitalism and racial capitalism as intersecting, as well as Fox & Miller-Idriss’ (2008) theories of everyday nationhood and everyday nationalism, the article explores the complex mechanisms at play in customers’ responses to retail stores’ social media advertisement posts addressing Ramadan. The overarching research question posed is: how can the online responses to social media advertisement addressing the Muslim custom of Ramadan be understood in relation to constructions of nationhood and anti-Muslim racism? The article untangles how the advertisements and the online responses to the advertisements follow a (market) inclusive logic (you are welcome to do your shopping here), as well as a highly exclusionary nationalist and racist logic. The latter is connected to co-customers’ negative reactions to the inclusive approach of the advertisements. This, we argue, needs to be understood in connection to anti-Muslim racism and nationalist tendencies in Swedish society more broadly (Muftee, 2023). The article makes visible a tension within racial capitalism, where profit-makers, though diversity strategies, try to enhance the market, while customers seek to uphold and guard racial hierarchies in the marketplace. In addition, and importantly, the article displays how customer responses are not uniform; instead, anti-Muslim expressions are contested, and the construction of Swedishness is negotiated and challenged.

Sweden is currently experiencing a political right turn; among other things, the government is seeking to drastically reduce the number of migrants who can enter to the country, increase repatriation of migrants, and introduce stricter conditions for family member immigration as well as stricter requirements for low-skilled labour immigration and for obtaining Swedish citizenship (Government Offices of Sweden, 2023). As in many other European countries, anti-Muslim racism has been central in legalizing this turn (Fekete, 2009). Racism against Muslims in Sweden is often articulated as a conflict of culture and values (Kundnani, 2023). As argued by Fekte (2009), since September 2001, Islam has been defined as the central threat to Europe, and Muslims are defined both as an 'enemy within' and a threat to Europeanness and Swedishness itself. While what people eat or put in their shopping basket might appear trivial issues, as will be discussed in the article, the debates related to Ramadan on the retail stores' social media platforms exposes the connection between consumption and larger questions of nationhood and race.

2. Racialisation in physical and digital consumer spaces

American 'shopping while black' literature shows that racial profiling and racial discrimination affect African Americans' experiences as consumers, and that anti-black bias manifests itself in retail settings; as a result, black consumers are forced to deal with racial hierarchies, which affects the shopping experience negatively (Pittman, 2020, see also Francis & Robertson, 2021). Bennett, Hill & Daddario (2015) in turn have found that different racial minority groups in America experienced similar levels of perceived marketplace discrimination. Research investigating Muslims' experiences in particular show that Muslim women wearing headscarves are exposed to gendered and Islamophobic violence in public spaces (Listerborn, 2015), experience perceived discrimination while participating in leisure activities in public places in the Netherlands (Kloek, Peters & Sijtsma, 2013), and develop different strategies to negotiate their 'Muslimness' and to handle 'anti-Muslim acts' in Paris (Najib & Hopkins, 2019). Muslim consumer experiences in retail settings are less studied, as pointed out by Alkayyali (2019) who conducted 20 in-depth interviews with veiled Muslim women in Paris. Among other things, the results show that the women experience harassment as well as 'bullying', and that co-customers have a central role in this treatment. Alkayyali's (2019) study shows that many expressions of racism in retail settings are blatant rather than subtle; Alkayyali underlines the importance of recognizing this tendency, as it may 're-become the norm for many racialised groups' (2019, p. 101).

Hussein (2015) analysed a social media 'scare' campaign against Halal-certified food from an Australian perspective and argues that there has been a shift in the racialisation of Muslims, maintaining that Muslims have gone from being portrayed as 'a visible, alien presence to a hidden, covert threat' (2015, p. 85). In the attacks against the campaign, Muslims are accused of being infiltrators and for 'blending in', for instance through the discreet presence of halal-certified foods in Australian shops. While 'ethnic foods' have largely become a central element of 'everyday multiculturalism', connected to cosmopolitanism and tolerance, halal certification of foods (for instance shown by product labels) does not speak to 'culinary multiculturalism' and is instead thought to have a hidden agenda (Hussein, 2015). Wright & Annes (2013) have explored how meanings of halal foods are contested in media discourse in France in relation to a fast-food chain introducing halal hamburgers on their menu. The responses contained some acceptance of the halal menu due to free-market logic or cultural diversity. But above all, the media engaged in a form of defensive 'gastronationalism', as they framed the halal hamburger menus as threats to French identity and the presumed core values of the French nation. In line with this, Nussbaum (2012) draws attention to bans of kebab shops in some Italian cities in 2009, purportedly due to health concerns and for the preservation of Italian food traditions.

Research primarily focusing on racialisation in online retail settings is lacking in a Swedish context; however, studies more broadly examining experiences of racism in Swedish society address consumer spaces to some extent. A study by Mulinari et al. (2024) that examined the prerequisites and obstacles for Roma life showed that shops are one sphere where Swedish Roma most frequently experience blatant antiziganism. In a study about (im)mobilities in public spaces among teenagers racialised as non-white in Stockholm and Malmö, Sixtensson & Hagström (2024) show that participants frequently experienced being subject to control and surveillance in shops and shopping malls (see also Kalonaityté et al., 2007). Moreover, Listerborn's (2015) study focusing on Muslim women's experiences in public spaces show among other things that the women experienced violent encounters in different retail settings. The article contributes to the understanding of racism in present-day Sweden. Specifically, it adds new knowledge about how racism, and particularly anti-Muslim racism, is constructed in digital Swedish consumer spaces. The following section will explore more closely the topic of the digital sphere in general, and social media platforms in particular, as a contested (racialised and nationalistic) consumer space.

3. Digital (marketing) logics, race and constructions of the nation

The representation of Islam and Muslims in social media is wide-ranging, however, social media users more often portray Islam and Muslims negatively than in a positive way (Hashmi, Rashid & Ahmad 2020). In a study about representations of Muslims in Swedish social media discourse, Törnberg & Törnberg (2016) show that Muslims are represented both as violent and extreme. Pointing towards the need to extensively examine hate crimes on social media platforms, Awan (2016) studied how Muslims are being viewed on the social media platform Facebook by analysing Facebook pages, posts, and comments. The study shows that Muslims are subjected to negative attitudes, stereotypes, discrimination, physical threats, and online harassment. Awan (2016) also discusses social media's lack of action against racial hatred. Obler (2016), in turn, examines the normalization of hate against Muslims through the use of the social media platform Facebook and maintains that online Islamophobia is a problem that social media companies need to take seriously and act upon. Matamoros-Fernández (2017) proposes the concept of 'platformed racism' to understand the particular forms of racism that derive from social media platforms and in the theoretical piece 'Where Platform Capitalism and Racial Capitalism Meet: The Sociology of Race and Racism in the Digital Society', McMillan Cottom (2020) argues that new theoretical frameworks are needed to study race and racism in the digital society (see also Daniels, 2015). To understand its specific logics, we need to turn to theories of racial capitalism, which captures the relationship between global and local processes, and how these intersect with platform capitalism. According to McMillan Cottom (2020, p. 444), platform capitalism, as a 'specific and current stage of capitalism' has the capacity to expand markets; in fact, internet technologies have become a major tool of capitalism because they *can* expand markets and consumer classes. However, platform capitalism also engages in predatory inclusion: 'the logic, organization, and technique of including marginalized consumer-citizens into ostensibly democratizing mobility schemes on extractive terms' (McMillan Cottom, 2020, p. 443), thus both expanding and excluding. According to McMillan Cottom (2020), Gargi Bhattacharyya's (2018) theories of racial capitalism are specifically suited for the study of race and racism in the digital society, since Bhattacharyya emphasizes how the logic of racial capitalism on the one hand works through the use of coercive power, and on the other also mobilizes desire, for instance to gain status or to feel belonging. Moreover, once again drawing on Bhattacharyya's thinking, McMillan Cottom (2020, p. 446) claims that platform capitalism has in turn 'monetized' all those human desires by 'capturing both space and place'. Unlike other theories that highlight the violent nature of race and racism, racial capitalism in the digital society appeals to our human desires, operating in a less obvious, but still highly effective way.

Marketers' strategies to reach out to certain assumed ethnic groups of customers, so-called 'ethnic marketing' (Licsandru & Cui, 2018, p. 330) or 'multicultural marketing' (Burton, 2002), aim to expand markets (Cui, 1997; Peñaloza, 2018). Ulver & Laurell (2020) examine online consumer resistance against multiculturalism in advertising in a Swedish context; they argue that far-right resistance is highly evident in these marketing contexts. According to Siddiqui & Singh (2016), social media functions as communication platforms that enable interaction, or even dialogue, between companies and their customers. They are used in different ways to attain business goals; for example, companies advertise on their social media platforms to attract customers. As Siddiqui & Singh (2016) argue, a positive effect of communicating with customers in this way is that social media interaction with customers may facilitate understanding of their desires and disapprovals. It also helps companies reach out to new customers. At

https://doi.org/10.33621/jdsr.v7i133259

the same time, business social media strategies may well also lead to negative effects for companies, for instance through negative comments and opinions posted by followers on the platforms.

Wei & Bunjun (2020) maintain that the subject of consumer nation-building in relation to branding is under-researched by critical race scholars, and they studied how consumers on Twitter respond to the attempts of the brand New Balance to distance itself from associations to white nationalism through claims of diversity. Three customer responses emerged in their material: punishing the brand; advising the brand; and defending the brand. Moreover, the digital responses were pronounced with, and through, 'circulation of affect' (such as indignation and hope) and connected to nation-building, as consumers positioned themselves as 'speakers of the nation':

> Analysis reveals that consumers are constructed and construct themselves within an elevated status as 'rightful' citizens and speakers of the nation, as of value and belonging to national spaces of discussion. In doing so, consumers position themselves as managers, who are willing and able to punish, advise, and defend, not just the brand but also the nation. (2020, p. 1271)

Sara Ahmed (2000) follows the path deriving from Anderson's work on nations as imagined (Anderson, 1983), describing the nation as a fantasy and a 'material effect'. The production of the nation, Ahmed argues, involves image and myth-making such as the reproduction of 'official' stories of descent, but also 'the everyday negotiations of what it means 'to be' that nation(ality)' (Ahmed, 2000, p. 98). The nation is thus both a place and a person – and individuals both *have* and *are* a nationality. Moreover, the nation comes into being through 'the recognition of strangers', which lets the nation 'imagine itself as heterogeneous'. This recognition of the strange and familiar, of who or what does or does not belong, takes place in everyday encounters but it is also part of 'rehearsed' public discourses of nationhood (Ahmed, 2000, p. 96-99). Fox & Miller-Idriss (2008) argue that nationhood and nationalism are produced and reproduced in everyday life, and that ordinary people are active in the production and reproduction of nations. Four practices of how this takes place are suggested: 1) 'Talking the nation: the routine construction of the nation through routine talk in interaction', which means that ordinary people help define discourses about the nation through talk and interaction in contexts of everyday life. 2) 'Choosing the nation' suggests that nationhood forms, and is formed by, people's choices. 3) 'Performing the nation' means that nationhood is given meaning in ritual and symbolic collective performances of everyday life. 4) 'Consuming the nation' refers to the way national difference and sameness are constructed and transmitted (and materialize) through everyday routine consumption habits. This can be understood as a 'commodification of the nation', where selected literature, music, food, or costumes offer people 'nationally marked (or markable) products for their national consumption needs' (Fox & Miller-Idriss, 2008, p. 551). As will be discussed further, advertisements connected to Ramadan on retail stores' social media platforms become an arena where conflicts around what the nation is unfold.

4. Method and material

This study is a non-interfering analysis of online archival data. Such data may be, as in this case, comments generated from public social media posts or videos (Kozinets, 2015). We have analysed written online responses to Swedish retail stores' social media advertisements that broadly address the Muslim celebration of Ramadan. The authors became aware of the phenomenon of retail stores either addressing or not addressing Ramadan through a news report and became interested in observing how this was reflected on their social media platforms. We began manually searching for posts related to Ramadan on the Facebook pages of the retail stores. The phenomenon of addressing Ramadan in retail stores' advertisements is not widespread, which meant that many searches did not yield any relevant posts. After identifying posts from 16 different stores, the decision was made that saturation had been reached. This decision was based both on the content and the quantity of comments generated by the posts. A total of 19 social media advertisement posts highlighting Ramadan that together generated 2988 responses in discussion threads have been included in the study. So-called 'likes' are not included. The advertisement posts originate from 16 unique local retail store chains with different geographical locations in Sweden, and one included post comes from a national retail store chain's social media website. 17 of the social media advertisement posts included originate from grocery store chains. The other two posts are from a retail chain that sells groceries as well as other goods. The selected advertisement posts mainly contain special offers for certain items but also may include holiday greetings wishing a happy celebration, photos of foods or other items on sale, and/or photos of staff holding up items or posters. The number of responses generated by the 19 posts varied (from eight to 1,600) and did not always contain negative remarks towards Muslims (14 posts contained negative remarks, five contained no negative remarks).

The advertisement posts included were published on the stores' social media websites between 2018 and 2023. The time frame was chosen to ensure the study's temporal relevance; however manual searches also revealed that posts were rare before 2018. The responses contain written comments, emojis, links and memes/pictures; only written comments have been included in the analysis, however. When collecting archival online material (Kozinets, 2015) originating from social media websites or forums, one must take into account the risk of including non-human generated comments. As the material was analysed manually and all comments on the social media platforms were linked to personal social media accounts, we saw no obvious indications of this being the case. Unlike in Ulver & Laurell's (2020) study, advertisement posts and responses included in this study originated in forums that are not known for attracting any specific group of people other than customers as such. We believe that this might further reduce the possibility of bot-generated comments or comments produced in so-called troll factories. Total certainty of this is impossible, however. Here we lean on Ulver & Laurell (2020, p. 481), who maintain that: 'inside the specific cultural context, it does not matter if some of these posts are artificial or created by, say, bots, because they are still repertoires that give meaning to the debate and in the end may have political consequences.'

We have applied a thematic content analysis to analyse the empirical material (Braun & Clarke, 2006). The material was coded manually and subsequently categorized into content-related categories. Prominent themes deriving from this work are presented in the article and analysed in dialogue with theories and relevant previous research. The analysis followed an abductive approach, where theory and empirical data continuously informed each other. Due to the iterative process, some overlap between the themes occurred. The excerpts presented in the article should be understood as representing patterns in the material. At times, the excerpts contain emojis, such as hearts, flowers or other symbols used by the commenters to emphasise the written message or to show some kind of emotion or reaction. The emojis have been preserved in order to remain close to the social media websites as an interactive phenomenon; other than that, they are given no analytical significance. As previously stated, the social media websites from which the data is collected are public and can be visited by anyone. The comments generated from the retail stores' advertisements are thus public, however, generally come from individual's private social media accounts, where most users use their own names. A major ethical concern has been preserving the anonymity of individuals whose comments are included in the study. To prevent extracts in the article from being traced back to any individual, we do not disclose the name of the retail companies and their social media platforms included in the study. We have also omitted any information that could be connected to the individuals behind the comments. Moreover, the translation of the comments from Swedish to English further adds to the non-traceability of the comments (Sylwander, 2019). The names of the retail stores and their geographical locations have been omitted from the article for the same reason. Throughout the analysis, we have chosen to refer to the individuals behind the comments on the social media platforms as customers, co-customers or commenters[1].

[1] *The research project has been given ethical clearance by the regional ethics board in Sweden (EPN 2022-00782-01).*

5. Analysis

5.1 Food, nation and anti-Muslimness

> Ramadan is coming up! We're celebrating this with super offers on halal beef and chicken. Welcome!

Sales on different foods are naturally often the centre of attention in the retail stores' advertisement posts, as exemplified above, where a local grocery store is offering discounts on halal meat. Foods and food practices are also common features in the responses to such advertisements and seem to be used symbolically to make different anti-Muslim statements, but also to manifest Swedishness. A food that is recurringly referred to in customer replies in such a way is bacon or more generally pork, a food customarily avoided by Muslims. These remarks often carry an ironic undertone, as in the following comments:

> I'll be celebrating it with lots of delicious bacon!
>
> No discount on bacon?
>
> Have you lowered the price on pork?

Besides making sarcastic, but yet implicit statements with the mention of pork or bacon, more direct opinions on halal practices and halal certified foods are central in the discussion threads. The negative comments related to halal range from aggressive: 'Great information, we other customers will avoid your store. Many of us don't accept halal!!! 🤮🤬', to more specifically pointing out that the practice is not Swedish: 'Advertising non-Swedish culture and halal food. I won't be coming back'. Comments such as the latter, drawing on (non-)Swedishness, recall Wright & Annes' (2013) findings, which show that the presence of halal food on a French fast-food menu was construed as a threat to both nation and nationality. Based on an analysis of a social media campaign against halal certification and -labelling of foods and products (i.e., not focused on slaughtering methods and animal welfare), Hussein (2015) in turn, found that the labelling of products as halal certified was perceived as a concealed threat, a way for Muslims and Muslim traditions to covertly infiltrate Australian society. Similar arguments are to some extent present in the material, as exemplified here:

> We consumers demand that halal certification is labelled in a highly visible way, so that we can avoid the products.

According to Ahmed (2000), discourses about the strange as threatening lead to constructions of the figure of the stranger as dangerous ('stranger danger') and a risk for the imagined sense of 'we'. According to this logic, the stranger is a necessary condition for upholding the imagined 'we'. Implicit in the argument that halal labels on food in retail stores should be highly visible seems to be a notion that foods and human bodies represent a similar threat: concealing them threatens the logic and the status quo between the strange(r) and the imagined 'we'. Halal slaughtering processes are also frequently targeted: 'Halal slaughter equals animal torture!'. Other commenters point out the paradoxes inherent in comments that raise the issue of animal rights, or claim to support Swedish foods:

> As soon as you read about halal, then all of a sudden you start to care about animal rights? You don't think about the well-being of animals when you buy your discount hot dogs. Can't believe people talk about us needing to buy Swedish foods and then go and buy a pineapple😂

One way in which people speak out against diverse forms of anti-Muslim racism in the material is by highlighting the inconsistency of arguments, often with a sense of humour, such as here questioning not only whether animal welfare is the issue, but also pointing out that people consume many things that are not ‘Swedish’. Besides featuring halal products, the foods that are featured in the advertisements vary to include for example rice, lentils, chicken, lamb, yoghurt, certain breads and pastries, greens and herbs such as mangold, parsley, and cilantro, tomatoes and onions, and fruits such as dates. While these products may typically be associated with so-called ‘Middle Eastern’ or ‘Mediterranean’ cuisine to an extent, they may also be considered part of standard stock in many Swedish grocery stores. Thus, even though these foods are part of an advertisement targeting Muslim customers, they are in fact special offers of which all customers can make use, for products they can eat or use in their cooking. Nonetheless, many perceive the offers as unfair, and as the below excerpt shows, a common argument is that the stores should make up for the advertisement and possible discounts by reducing the prices of ‘Swedish foods’, especially in relation to national holidays:

> Okay, so then we expect discounts on Swedish holidays. On typically Swedish foods.
>
> As long as we get discounts around our Easter. You wouldn’t want to be considered unfair, right?

The discussion threads on the topics of food and food practices following the advertisements are characterized by negotiations rather than expressing uniform views. Many speak up or rebuke negative comments, as shown in the two following statements:

> What a nice initiative. I am a Christian and happy about this inclusive way of thinking, where there is space for food traditions from all around the world.
>
> I realize there is a lot of ignorance in the comments, people are raging because [the store’s name] includes other cultures. It’s nice of them. People say they want to boycott, like it’s the first time they’re selling halal meat. That’s so narrow-minded. In that case you’ll have to boycott all other food stores as well. That’s all from me. Now, I’m going to prepare my halal slaughtered chicken. 🍗

The first commenter here highlights their Christian religious views and their appreciation of the inclusive initiative of food traditions. The second draws attention to halal foods being widespread in Swedish stores – and points out that the phenomenon is not a new one. These comments demonstrate that constructions of difference and sameness manifested on the social media platforms are contested. Still, the display of Muslim food traditions and practices on the retail stores’ social media pages seem to trigger self-appointed ‘speakers of the nation’ (Wei & Bunjun, 2020; see also Hussein, 2015). The production of the nation involves everyday negotiations of what it means ‘to be’ a nation and a nationality (Ahmed, 2000). Similarly, Fox & Miller-Idriss (2008) argue that nationalism is an act of production, and some products are constructed as national products more than others. This, they argue, is the ‘commodification of the nation’. Food is one such product that ‘defines, demonstrates, and affirms the consumer’s national affinities’ (Fox & Miller-Idriss, 2008: 551). As evident in our study, previous research contributions have shown the symbolic significance of food or food traditions to protect constructed national values, particularly in relation to halal products – a form of ‘gastronationalism’, as Wright & Annes (2013) put it; there have also been ‘scare’ campaigns against halal certification (Hussein, 2015). In the phrasing of Hussein (2015, p. 93), this sends out a message to Muslims: ‘that, however discreete their presence, however well integrated they may believe themselves to be, they are not welcome here’ (Hussein, 2015,

p. 93). It is notable that even though food is considered an important product of the 'commodification of the nation' (Fox & Miller-Idriss, 2008), the retail stores, unintentionally or not, seem to challenge the boundaries of such imagined national products.

5.2 Contested (Swedish) traditions

> We have everything you need for Ramadan! Eid Mubarak from all of us!

Many of the retail stores' advertisement posts not only inform about special offers for Ramadan but also wish those observing it happy celebrations, as exemplified in the extract. One major point of criticism in the discussion threads that follow such advertisement posts is that non-Swedish traditions should not be 'celebrated' in Sweden, and the retail stores should not acknowledge such traditions or cultures.

> Aren't Swedish traditions good enough? Usually, if you move to a new country, you adapt to its traditions.
>
> We live in Sweden and our culture doesn't celebrate Eid or al-Fitr. No other country in the world would acknowledge Swedish culture or Swedish traditions. It is ridiculous that [the store's name] includes Eid food in its range. So stupid!

As the two posts show, such comments address both individuals who practice non-Swedish traditions and the retail stores that recognize Muslim traditions and enable Muslim celebrations. Moreover, Swedish traditional celebrations such as Christmas, Easter and Midsummer Eve are frequently referred to in the comments by customers who want to highlight what the retail stores *should* address and celebrate. We see the recurrent references to 'Swedishness', Swedish habits and Swedish traditions in the material analysed largely as examples of how nationhood and nationalism are constructed, manifested and reproduced in everyday and everyday interactions (Fox & Miller-Idriss, 2008), which here take the form of written statements and interactions on the retail stores' social media platforms. However, Fox & Miller-Idriss (2008) also maintain that nationhood is formed and given meaning by performing rituals and traditions, which further can explain the frequent mention of Swedish traditions. Commenters manifesting national values seem to do so to guard 'Swedishness', among other things pointing out that Muslims need to adapt, not the other way around. This is reminiscent of Wei & Bunjun's (2020) findings regarding consumers who act as 'speakers of the nation'. In parallel, the retail stores are assigned positions where they are seen as bearers of Swedish culture and traditions. Posting advertisements that acknowledge Ramadan thus appears as a violation of Swedish traditions and culture, a sort of non-performance of nationhood, to use the vocabulary of Fox & Miller-Idriss (2008).

The recognition of 'strangers' is central in the construction of the nation, as this helps the nation to stay close to the fantasy of it being heterogenous (Ahmed, 2000). In that sense, the 'stranger' plays an important role in the ongoing mythmaking of the nation as a place and a person. While some consumers can be understood as positioning themselves as 'speakers of the nation' (Wei & Bunjun, 2020), in that they express which traditions the retail stores should and should not support, the analysis also shows that the comments deriving from the retail stores' addressing Ramadan result in directly anti-Muslim racist views that are directed towards Muslims as a group, as exemplified here:

> Just go back to where you came from.
>
> People mention that comments are filled with hate and ignorance, it's kind of funny, since that's just what Muslims are filled with.

Such statements are in line with previous research that has shown how Muslims, especially women wearing veils, are subjected to different anti-Muslim expressions in western European settings (Listerborn, 2015; Kloek, Peters & Sijtsma, 2013; Alkayyali, 2019; Muftee, 2023). Other studies show that Islamophobia is normalized on social media platforms (Awan, 2016; Obler, 2016). In our analysis, the retail stores' advertisements relating to Muslim customs and traditions, rather than the physical appearance of bodies racialised as Muslims, seem to trigger racist statements. The analysis shows that such comments, as exemplified in the excerpts, are sometimes directed to a specific commenter in response to another comment, and sometimes they are just statements without a particular recipient in the discussion threads, speaking instead to 'all Muslims'. An important note regarding to face-to-face encounters is that the scope of potential recipients of anti-Muslim racist statements might be far wider on a public social media platform, since anyone reading the comments is a potential recipient. Moreover, correlating with Alkayyali's (2019) findings of Muslim women's experiences of shopping in physical environments, many anti-Muslim racist comments in our analysis come across as highly blatant.

Anti-Muslim remarks are not left unchallenged; they are interspersed with comments by consumers who view the retail stores' advertisements in a positive, appreciative or supportive way:

> We Swedes need to be more open-minded about other traditions. If I lived abroad and continued to observe Swedish traditions, no one would care.
>
> I wonder why some write that retail stores should not make religious statements? What about Christmas? We should be happy for all people's holidays, not just our own. So many Muslims wish me Merry Christmas, I want it to be mutual, we should be happy for each other.
>
> Thank you for having the courage to acknowledge a tradition that is celebrated by thousands of Swedes, that you unlike many others don't bow to hatred and Islamophobia. Respect.

Thus, consumers also voice objections to intolerant expressions in the comments. Arguments contain for instance responses to claims about maintaining 'Swedish traditions' in favour of more open-minded approaches to traditions and religious views or holidays, as in the first and second excerpts, also expressing support of everyone's right to celebrate their holidays. A recurring argument against the retail stores' acknowledgment of Ramadan is the thought that the retail stores should maintain religious neutrality; the consumer in the second excerpt objects, drawing attention to the role of Christmas in advertisements. The third excerpt shows support and appreciation of the retail stores' initiative and emphasizes the fact that Ramadan actually is celebrated by a large number of Swedes. Thus, negative views about Muslims as such and critical comments about the retail stores' inclusion of the Muslim group as customers through the advertisement are challenged. Moreover, constructions of nationhood, here primarily centring around traditions and habits (Fox & Miller-Idriss, 2008) and what it is 'to be' Swedish (Ahmed, 2000), that occur in interactions on the social media websites following the retail stores' initiatives also seem to be under negotiation, rather than fixed.

5.3 A digital space of contradictions

This final analytical theme will focus on what has been touched upon in part in the two previous analytical sections: the mix of inclusive and positive comments on the one hand and spiteful, negative comments against Muslims, Muslim traditions, and/or the retail stores' practice of posting advertisements related to Ramadan on the other. We will provide examples of such ambiguities and also discuss how the social media platforms may be seen as a facilitator of Islamophobia (Awan, 2016; Obler, 2016), as well as a space where such views might be challenged in different ways.

Negative comments are not only directed at Muslim customers, but also at the retail companies. Such comments describe the retail stores as money-driven, as political actors, or as performing shameful or morally wrong activities:

> You'd do literally anything to make money.
>
> Is business bad? It's wrong to bring religion into advertisement.
>
> For money, you'll sell yourself to anyone.
>
> It's shameful to see how far you go to make a profit.

Such comments are in line with Wei & Bunjun (2020), who argue that consumers who appoint themselves as 'speakers of the nation' also position themselves as having the right and ability to punish or lecture companies that make appeals to diversity or go against their perceived national values. In the extracts exemplifying this, the 'speakers' criticize the retail stores for prioritizing profit over gate-keeping the nation. Threats of boycotts are frequent:

> I hope people boycott you!
>
> Disgusting advertisement! Total boycott!
>
> Well, now you've chosen your segment of customers. Good luck!
>
> And with that I stopped doing my shopping at this place.

The alleged promises of boycott come across as aggressive and threatful. The statements indicate the potential risk-taking involved in the retail stores' initiatives on their social media platforms (cf Siddiqui & Singh, 2016). However, the analysis also shows that the retail stores might expect new customers who appreciate the gesture; such comments are many. As can be seen in the following excerpt, such commenters express not only gratitude for the customer inclusion, but also for the retail stores taking the initiative despite the possible risk of losing other customers:

> Thank you for including Swedish Muslims despite hateful Islamophobia, we will support you in good and bad🌹

There are also comments that point out the logic of the capitalist market in the analysed material. Such responses normally reply to negative comments, arguing that is only logical that the retail stores include this large group of potential customers:

> The store is a business that wants customers. It's as simple as that. A clever person would realize that it is possible to attract customers and make a profit if you acknowledge events in their lives. Since people are different, it is wise to include all kinds of events, regardless of whether they are connected to religion or something else. […] Why should retail stores give up customers because they are celebrating other holidays than the majority of society does?

As seen from this extract, this commenter seems to want to enlighten other commenters who are critical of the retail stores' inclusion of Muslim traditions and draws a parallel between differences and profit. The analysis furthermore shows that the (market) inclusion of this group as customers partly gives rise to highly appreciative and grateful comments directed at the retail stores about being seen, thought of, and included. The appreciative comments go beyond being included as customers and emphasize a thankfulness for being included as members of society, or as part of, and belonging to, the 'Swedish community', as exemplified in the following two excerpts:

> Thanks for the inclusion, it means a lot, especially for younger generations who can proudly feel that they belong in this beautiful country.
>
> Very nice that you include us in your community❤️🌹, it warms the heart.

The practice of addressing Ramadan on the retail stores' social media websites/platforms seems to have a symbolic value that goes beyond a customer/company relationship, also representing a *promise* of acceptance and inclusion in Sweden and becoming Swedish (Ahmed, 2000). Thus, at the same time as the interaction between commenters on the platforms can be seen as transmitting both nationalist and racist discourses, the global logic of the capitalist market is also at play, creating a kind of paradoxical phenomenon. On the one hand, the advertisement itself transmits a message of inclusion of Muslim customers, which in the comments generates expressions of inclusion and belonging. On the other hand, the retail stores' (market) inclusion leads to aggressive comments from co-customers who question this 'inclusive' practice. Here, we wish to draw parallels to McMillan Cottom's (2020) sociology of race and racism in the digital society and the connection to racial capitalism. One key argument made by McMillan Cottom (2020) is that racial capitalism in the digital society talks to human desires, for instance, to consume, or to belong. A characteristic of racial capitalism in digital society is that it engages in 'predatory inclusion', a sort of inclusion by exclusion, a practice that only seemingly includes marginalized consumers. In this study, we have focused on the reactions generated by the retail stores' advertisements, but we have not analysed the retail stores' intentions or motives or people's lived experiences of this customer market inclusion. Thus, inspired by McMillan Cottom's (2020) thoughts on predatory inclusion, we choose to describe the phenomenon that occurs when the retail stores address Muslim celebrations and traditions in their social media advertisement as a form of paradoxical inclusion where marginalized consumer-citizens are targeted as both new customers and as deviant – a form of inclusion AND exclusion. Moreover, the retail stores' social media platforms not only are spaces of hatred against Muslims as a group; they also seem to be a space where resistance against anti-Muslim racism is articulated, and where constructions of Swedishness are challenged.

6. Discussion

The article shows, from a 'bottom-up' perspective, how everyday nationhood and nationalism (Fox & Miller-Idriss, 2008) is constructed and reproduced on the digital platforms that serve as a link between retail stores and their customers. The advertisements and the responses to them, which display dislike as well as support for the retailers and the Muslim community, illustrate a negotiation of nationhood which is characterized on the one hand by racist anger and fear of loss of nation, and on the other by support for inclusion and expansion of the market. On the retail stores' social media platforms, Muslim consumers are targeted as both new and welcome customers and as deviant – a form of paradoxical inclusion.

Anti-Muslim comments tend to be overt in character. We suggest that this might be connected at least at part to the written format of commenting that is built into the system of social media platforms, but also to the normalization of racialised hate on social media platforms (Awan, 2016; Obler, 2016). The racist anti-Muslim comments mirror current nationalist tendencies and political cultural conflicts within Swedish society at large. There is an obvious danger that racist and anti-Muslim comments will contribute to further stigmatization of Muslims in Sweden today, both on social media and elsewhere. However, the findings of our study also shows that negative comments are challenged, for instance via expressed support for a retail store's initiative or everyone's right to their traditions. In that sense, the social media platforms are a space in which both racist and nationalistic views and resistance to such views can be communicated. Such findings highlight the importance of exploring the digital arena of consumption as a contested space where conflicts over the meaning of what the nation is and who can belong to it are articulated.

Occasionally, the retail stores post responses to negative comments in the discussion threads; a few of these defend the advertisement or remark on the tone in the discussions. Such comments are infrequent, and we have thus chosen not to include them in the analysis. It is noteworthy however that the retail stores' comments are few, despite sometimes very long discussion threads containing anti-Muslim statements. Nor have the retail stores (or the social media facilitators Facebook or Instagram) shut down discussions, although they have the power to do so. Despite the retail stores' inclusive practice toward Muslim customers, the responsibility to this customer group seems to have its limits. This exposes the market forces that are inevitably in the background and may be related to McMillan Cottom's (2020) thoughts on how the capitalist logic of platforms entails that they not only have the ability to expand markets, but that they also engage in 'predatory inclusion', where the inclusion of marginalized citizens might be deceitful.

In summary, this article furthers knowledge of race and racism in the digital society as well as on racialisation in relation to consumption and how everyday nationhood and nationalism are reproduced and negotiated in digital consumer (social media) spaces. Research that investigates local retail stores' perspectives on including marginalized customers via advertisements on their social media platforms would add important knowledge to further understand the processes at play. In a similar way, research on marginalized customers' lived experiences of digital consumer (social media) spaces would be an important future research contribution.

Disclosure statement

The authors report there are no competing interests to declare.

Funding

This work was supported by the Swedish Research Council (Vetenskapsrådet) under Grant 2019-03303.

References

Ahmed, S. (2000). Strange Encounters: Embodied others in post-coloniality. Routledge.

Alkayyali, R. (2019). Shopping While Veiled: An Exploration of the Experiences of Veiled Muslim Consumers in France. In Johnson, G., Thomas, K., Harrison, A., Grier, S. (eds), Race in the Marketplace (pp. 89-105). Palgrave Macmillan, Cham. DOI: 10.1007/978-3-030-11711-5_6.

Anderson, B. (1983). Imagined Communities: Reflections on the origin and spread of nationalism. Verso.

Awan, I. (2016). Islamophobia on Social Media: A qualitative analysis of the Facebook's walls of hate. International Journal of Cyber Criminology, 10(1), 1-20. DOI: 10.5281/zenodo.58517.

Bennett, A.M., Hill, R.P. & Daddario, K. (2015). Shopping While Nonwhite: Racial Discrimination among Minority Consumers. Journal of Consumer Affairs, 49(2), 328-355. DOI: 10.1111/joca.12060.

Bhattacharyya, G. (2018). Rethinking Racial Capitalism: Questions of reproduction and survival. Rowman & Littlefield International, Ltd.

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology,* 3:2, 77-101, DOI: 10.1191/1478088706qp063oa.

Burton, D. (2002). Towards a Critical Multicultural Marketing Theory. Marketing Theory, 2(2), 207-236. DOI: 10.1177/147059310222004.

Cui, G. (1997). Marketing Strategies in a Multi-Ethnic Environment. Journal of Marketing Theory and Practice, 5(1), 122-134. DOI: 10.1080/10696679.1997.11501756.

Daniels, J. (2015). "My Brain Database Doesn't See Skin Color": Color-Blind Racism in the Technology Industry and in Theorizing the Web. American Behavioral Scientist, 59(11), 1377–1393. DOI: 10.1177/0002764215578728.

Francis, J.N.P. & Robertson, J.T.F. (2021). White Spaces: How marketing actors (re)produce marketplace inequities for Black consumers. Journal of Marketing Management, 37(1-2), 284–116. DOI: 10.1080/0267257X.2020.1863447.

Fekete, L. (2009). A Suitable Enemy: Racism, migration and Islamophobia in Europe. Pluto.

Fox, J. E. & Miller-Idriss, C. (2008). Everyday Nationhood. Ethnicities, 8(4), 536-563. DOI: 10.1177/1468796808088925.

Government Offices of Sweden (2023). Sweden's new migration policy. Available at: https://www.government.se/government-policy/swedens-new-migration-policy/. (Accessed 11 December 2023).

Hashmi, U. M., Rashid, R. A. & Ahmad, M. K. (2020). The representation of Islam within social media: a systematic review. Information, Communication & Society, 24(13), 1962–1981. DOI:10.1080/1369118X.2020.1847165.

Hussein, S. (2015). Not Eating the Muslim Other: Halal certification, scaremongering, and the racialisation of Muslim identity. International Journal for Crime, Justice and Social Democracy, 4(3), 85-96. DOI:10.3316/informit.252514045117578.

Kalonaityté, V., Kawesa, V. & Tedros, A. (2007). Att färgas i Sverige: upplevelser av diskriminering och rasism bland ungdomar med afrikansk bakgrund i Sverige [To be colored in Sweden: experiences of discrimination and racism among youth of African descent]. Stockholm: Ombudsmannen mot etnisk diskriminering.

Kozinets, R.V. (2015). Netnography: Redefined. (2nd ed.). Sage Publications Ltd.

Licsandru, T.C. & Cui, C.C. (2018). Subjective Social Inclusion: A conceptual critique for socially inclusive marketing. Journal of Business Research, 82, 330-339. DOI:10.1016/j.jbusres.2017.08.036

Listerborn, C. (2015). Geographies of the Veil: Violent encounters in urban public spaces in Malmö, Sweden. Social & Cultural Geography, 16 (1), 95-115. DOI:10.1080/14649365.2014.950690.

Kloek, M.E., Peters, K. & Sijtsma, M. (2013). How Muslim Women in The Netherlands Negotiate Discrimination During Leisure Activities. Leisure Sciences, 35(5), 405–421. Doi:10.1080/01490400.2013.831285.

Kundnani, A. (2023). What is Antiracism?: And why it means anticapitalism. Verso.

McMillan Cottom, T. (2020). Where Platform Capitalism and Racial Capitalism Meet: The Sociology of Race and Racism in the Digital Society. Sociology of Race and Ethnicity, 6(4), 441-449. DOI: 10.1177/2332649220949473.

Matamoros-Fernández, A. (2017). Platformed racism: the mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube. Information, Communication & Society: 20(6), 930–946. DOI:10.1080/1369118X.2017.1293130.

Muftee, M. (2023). Navigating and Countering Everyday Antimuslim Racism: The Case of Muslim Women in Sweden. Critical Sociology, 49(7-8), 1251-1267. DOI:10.1177/08969205231158496.

Mulinari, P., Ali, A., Lindkvist, S. & Halilovic, M. (2024). Lycklig är den som har en öppen väg framför sig: en rapport om förutsättningar och hinder för romskt liv i Malmö [Happy are those who have an open road ahead: a report on the conditions and obstacles for Roma life in Malmö]. Malmö: Malmö stad.

Najib, K. & Hopkins, P. (2019). Veiled Muslim Women's Strategies in Response to Islamophobia in Paris. Political Geography, 73, 103-111. DOI:10.1016/j.polgeo.2019.05.005.

Nussbaum, M.C. (2012). The New Religious Intolerance: Overcoming the politics of fear in an anxious age. Harvard University Press.

Obler, A. (2016). The Normalisation of Islamophobia through Social Media: Facebook. In Awan I (ed) Islamophobia in Cyberspace: Hate Crimes Go Viral. Routledge.

Peñaloza, L. (2018). Ethnic Marketing Practice and Research at the Intersection of Market and Social Development: A macro study of the past and present, with a look to the future. Journal of Business Research, 82, 273-280. DOI:10.1016/j.jbusres.2017.06.024.

Pittman, C. (2020). "Shopping while Black": Black consumers' management of racial stigma and racial profiling in retail settings. Journal of Consumer Culture, 20(1), 3-22. DOI:10.1177/1469540517717777.

Rydström, K. (2024). Unpacking Online Retailing: The Organization of Warehouse Work and Inequality. Doctoral Thesis: Luleå University of Technology.

Siddiqui, S. & Singh, T. (2016). Social Media its Impact with Positive and Negative Aspects. International Journal of Computer Applications Technology and Research, 5(2), 71-75.

Sixtensson, J. & Hagström, M. (2024). Risk, Discomfort and Disruption: Experiences of (im)mobilities in public spaces among Swedish youth racialised as non-white. Journal of Youth Studies, 1-16. DOI: 10.1080/13676261.2024.2359087.

Sylwander, K.R. (2019). Affective Atmospheres of Sexualized Hate Among Youth Online: A contribution to bullying and cyberbullying research on social atmosphere. International Journal of Bullying Prevention, 1(4), 269-284. DOI:10.1007/s42380-019-00044-4.

Törnberg, A. & Törnberg, P. (2016). Muslims in social media discourse: Combining topic modeling and critical discourse analysis. Discourse, Context & Media, 13, 132-142.

Ulver, S. & Laurell, C. (2020). Political Ideology in Consumer Resistance: Analyzing far-right opposition to multicultural marketing. Journal of Public Policy & Marketing, 39(4), 477-493. DOI:10.1177/0743915620947083.

Wei, M.L. & Bunjun, B. (2020). 'We Are Not the Shoes of White Supremacists': A critical race perspective of consumer responses to brand attempts at countering racist associations. Journal of Marketing Management, 36(13-14), 1252-1279. DOI:10.1080/0267257X.2020.1806907.

Wright, W. & Annes, A. (2013). Halal on the Menu?: Contested food politics and French identity in fast-food. Journal of Rural Studies, 32, 388-399. DOI:10.1016/j.jrurstud.2013.08.001.

https://doi.org/10.33621/jdsr.v7i133259

JDSR 7(1) (2025) 35-45

JOURNAL OF DIGITAL SOCIAL RESEARCH

10.33621/jdsr.v7i154875

Imagining communities with 'intelligent' machines

Innovationism and the hope for alternative imagination[1]

Katja Valaskivi

University of Helsinki, Finland

✉ katja.valaskivi@helsinki.fi

Abstract

Shared perceptions of the world are imagined with and within available media technological environments. In other words our communication environment conditions our social imagination and the ways in which we can see the world. The essay, based on the inaugural lecture of the author, discusses how this conditioning takes place and with what consequences in the contemporary digital societies. The essay draws on the research by the author on innovationism and discusses the concepts of reversed tools, content confusion and attention factory. Utilizing the study by Berg & Valaskivi (2023) on commercial image recognition services and their performance in recognizing religion in images as an example, the essay illustrates failures and imperfections of AI technologies which are often considered more neutral than human beings. The essay calls for critical thinking on digitalization and expansion of AI technologies and encourages prioritization of humane interests as well as social and cultural welbeing over commerciality in technological development.

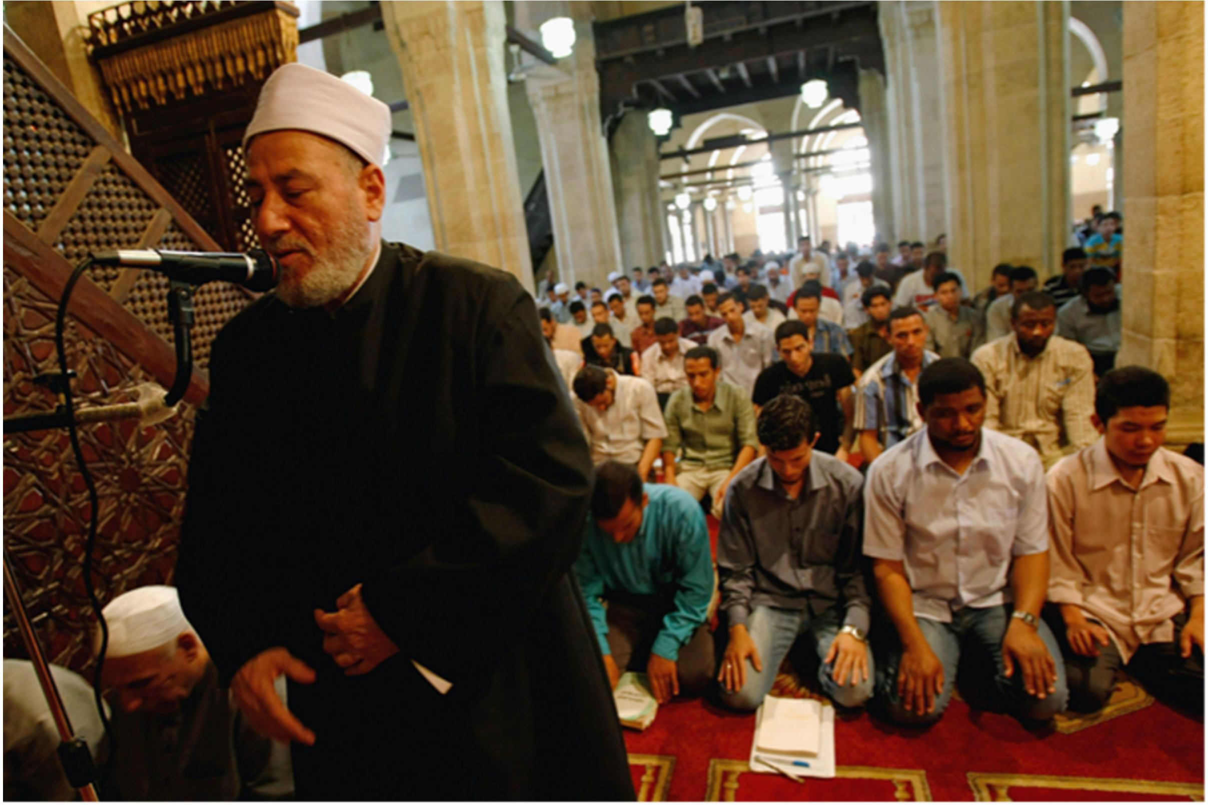
Keywords: innovationism, digitalization, image recognition, racial bias, attention economy

1. Introduction

Were you asked what you see in this picture (Fig. 1), you'd probably say something along the lines of 'Muslims praying in a mosque', or possibly 'an imam' or 'men practising religion'.

[1] This essay is a revised version of the professorial inaugural lecture of the writer on May 31, 2023 at the University of Helsinki, Finland.

Amazon Rekognition: Person (99%), Human (99%), Clothing (93%), Apparel (93%), Microphone (91%), Electrical device (91%), Crowd (82%), Hat (70%), Cap (62%), Audience (61%), People (61%), Fashion (59%).



Google Cloud Vision: Temple (88%), Public space (81%), Adaptation (79%), Event (75%), Human settlement (73%), City (70%), Crowd (69%), Religious institute (68%), Pilgrimage (65%), Cap (64%).

Figure 1. In the photo we can see an imam and men in a mosque, but this is not what the image recognition services "see". *Photo by David Silverman/Getty Images*

The world's leading tech companies' image recognition services would take a different view. Amazon Rekognition is confident that the picture shows a person (99%) and a human (99%), while Google Cloud Vision is slightly less confident that it shows a temple (88%) and a public space (81%). Neither of them possesses the lexicon to refer to an imam, mosque or even prayer – although Google Cloud Vision's vocabulary does feature the curious term 'religious institute' (68%).

This example is from the study we have conducted with Anton Berg (Berg & Valaskivi, 2023a; Berg & Valaskivi, 2023b). The research focuses on how how commercial image recognition services categorize images of religious subjects. This study will be explored in more detail below.

Here the example is used as an illustration of the ways in which contemporary AI applications take part in meaning making together with human beings and for human beings. These AI tools are in everyday use, and literally in our pockets and yet often invisible. An average smart phone user might enjoy the gadget being able to recognize and label their photos and organize them into daily video reels of friends, and be amused of an occasional miscategorization, without being aware of the technology behind the feature or think of the implications it might have on larger scale.

Another mundane example of daily AI in our lives is the AI Design feature of Microsoft PowerPoint. And with growing availability of content generating AI services and applications, the everyday signifying practices (Hall, 1997) have entered a new phase of transformation.

2. 'The Medium is the Message'

The principle that 'the medium is the message' was coined in the late 1960s by Marshall McLuhan, Canadian professor of English literature and the pioneer of contemporary media studies. Among McLuhan's (1967) ideas, the notion most relevant here is that all media are 'extensions of man', and that they create new environments for people to operate in. To understand social and cultural change, it is necessary to understand how and what kind of environment is created by media technologies (McLuhan, 1969).

In his classic book, Imagined Communities, first published in 1983, anthropologist and scholar of nationalism, Professor Benedict Anderson describes how print capitalism made possible the imagination of nation-states and the idea of nationalism. Although the members of an imagined community will never all get to know each other, they feel a sense of togetherness because of their shared cultural understandings. Collective imagination always takes place with and through available communication technologies and conditioned by those technologies. The key here to recognize, however, is that Anderson's emphasis was not on the technology, he did not refer to the printing press, but print *capitalism*. In other words, the societal role and impact of communication technologies is formed in the ways they are implemented and become integral parts of social practices and institutions and everyday life, and interaction.

Having established that the medium is the message, and the possibility of imagining community depends on media technologies and the social application of these technologies, we arrive at questions not only interesting for scholars working in the field of study of religion and media research, but that also have great importance in our society today:

First, how is it possible to imagine a sense of community in today's capitalized commercial media environment, where imagining happens through and together with various kinds of automated systems and more recently with content-producing AI systems?

Second, how do worldviews, beliefs, and ideologies, that is, different ways of imagining belonging and exclusion change in this media environment?

And third and finally, how do shared perceptions of media and communication technologies affect the processes of imagining belonging?

3. Reversed tools and content confusion in the attention factory

Figure 2. "The network in the middle of the village"
Illustration by Liekki Valaskivi 2016

https://doi.org/10.33621/jdsr.v7i154875

The Finnish saying “Kirkko keskellä kylää” or “The church in the middle of the village” used to describe how the central node of collective imagination used to be the church, both as a physical location and an institution of belief, belonging and meaning making. The church is still there but has become only one of the nodes among many in the contemporary signification and communication environment, which is a global, cross-border network. The development of the contemporary communication environment has, over the past 30 years, been driven by technological development and market concerns. In other words, new technologies have been adopted based on considerations of profit margins and business logics, not on whether and how they might contribute to building trust in a pluralistic society.

The change has been profound, because:

Firstly, the relationship between the production and reception has dissolved. In other words, while knowledge production has become democratized and opportunities for citizen participation have improved, it has become impossible for epistemic institutions such as the church, school, university, political parties – or journalism! – to control the production of meanings in the public domain (e.g. Peterson, 2003).

Secondly, the digital media environment is transnational in ways that inevitably have consequences for perceptions of nation, communities, and differences. Contents and meanings circulate across borders and platforms with unprecedented speed and volume.

Thirdly, the world is imagined not only by human beings but also by *reversed tools* (Couldry & Hepp, 2016; Valaskivi, 2022) that use us while we use them. If a hammer lands on your finger rather than the head of the nail, you can only blame yourself. The reversed tools of our digital communication environment, on the other hand, are constantly collecting data on users’ actions, creating profiles for profit, and curating content, further steering their actions. The reversed tools and different digital, algorithmic platforms make the media environment, which I have elsewhere described as *the attention factory* (Valaskivi, 2022). The attention factory prioritizes content that incite reactions, because reactions measurable for the reversed tools generate profits and data for the media platforms. As attention is in limited supply but media content is practically unlimited, attention is a valuable commodity. Therefore, it’s hardly provocative to say that the attention factory works on the principle of provocation.

People are more likely to pay attention and respond to affective subjects. The most affective subjects relate to basic values and belonging. This is why the most valuable and profitable commodities in this media environment are contents that deal with identity, religion, worldviews, and ideologies. If any of these trigger conflicts and confrontations then so much the better because feelings of threat, fear, hatred, and disbelief are most likely to provoke both advocates and opponents and therefore to elicit responses (Valaskivi, 2022; Kannasto *et al.,* 2023), to generate more data – and to bring in more profits.

Both media users and media platforms are therefore keen to try and trigger quick, affective reactions in users. Quick reactions are most likely to produce measurable responses: clicks, shares, comments, and reactions – some reversed tools even interpret a pause in scrolling through the content flow as a response.

This is how human emotions are commodified.

All this results in a cacophony of contents, or in the words of Professor Mara Einstein (2016), scholar in critical marketing studies and religion, *content confusion*.

In a media environment marked by content confusion, it is very hard, if not impossible, to know for what purpose a specific content has been produced: a news item can be a lie, a prank, a provocation, propaganda or possibly an advertisement. In other words, it may be impossible to know who, and with what motivation was behind the production of which content. And as social media contents reach us via various routes, through multiple shares and with comments and likes attached, it requires exceptional effort to find out the nature of the original content.

Reports indicate that utilization of the features of the attention factory, with its reversed tools, content confusion, and provocations, contributed to Brexit (Briant, 2018a; Briant, 2018b), the election of Donald Trump as US President in 2016 (CLC, 2020) as well as in the events of 6 January 2021 (Donovan & Dreyfus, 2022) when rioters stormed the US Congress in Washington DC.

4. Image recognition 'imagining' religion

I will now return in more detail to our research on image recognition services' abilities in recognizing religion. (Berg & Valaskivi, 2023a, 2023b)

As noted, our research interest was to study the ways in which image recognition services 'see' religion, in other words, how they categorize images that feature religious content. In what follows I will briefly explain some of our main findings. Unfortunately, it is not possible here to go into details of how image recognition services function.

From a user's perspective, image recognition services give categories to images by producing tags or labels and giving a confidence score in percentages to each of these tags. The score indicates the probability in which the image represents the category in question. For instance, in the case Fig. 3, Microsoft's image recognition service Azure is 84 per cent confident that the image features fashion accessories.

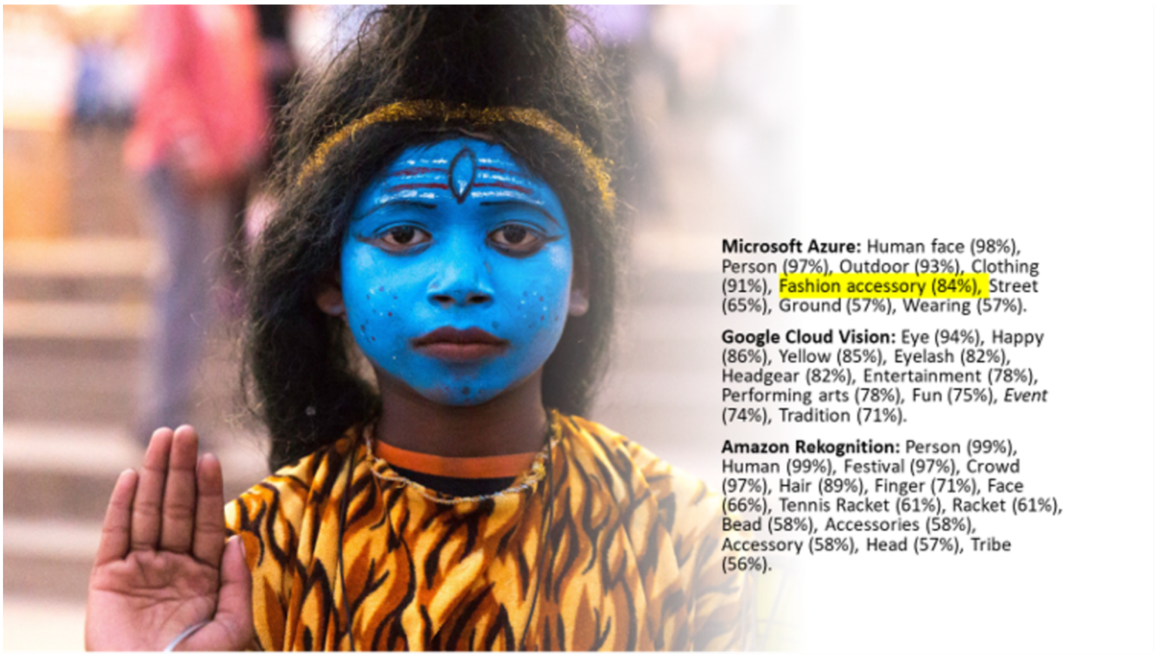

Microsoft Azure: Human face (98%), Person (97%), Outdoor (93%), Clothing (91%), Fashion accessory (84%), Street (65%), Ground (57%), Wearing (57%).
Google Cloud Vision: Eye (94%), Happy (86%), Yellow (85%), Eyelash (82%), Headgear (82%), Entertainment (78%), Performing arts (78%), Fun (75%), *Event* (74%), Tradition (71%).
Amazon Rekognition: Person (99%), Human (99%), Festival (97%), Crowd (97%), Hair (89%), Finger (71%), Face (66%), Tennis Racket (61%), Racket (61%), Bead (58%), Accessories (58%), Accessory (58%), Head (57%), Tribe (56%).

Figure 3. The image represents a child praying by the Ganges river. Image recognition services reproduce secular categories referring to entertainment, fashion and performing arts. *iStock photos*

The research process began with Anton Berg compiling the data from Google Images, a search engine for the retrieval of images. Our search terms were connected to different religious traditions and rituals. The complete dataset of 1189 images was fed into three image recognition services: Amazon Rekognition, Microsoft Azure and Google Cloud Vision through their application programming interfaces (API). The classifications and confidence score percentages were collected and analysed using qualitative methods.

The three services together assigned a total of almost 9100 classification tags to the images. Only 85 of these categories were related to religion; and as many of 30 of these 85 tags were related to Christianity.

https://doi.org/10.33621/jdsr.v7i154875

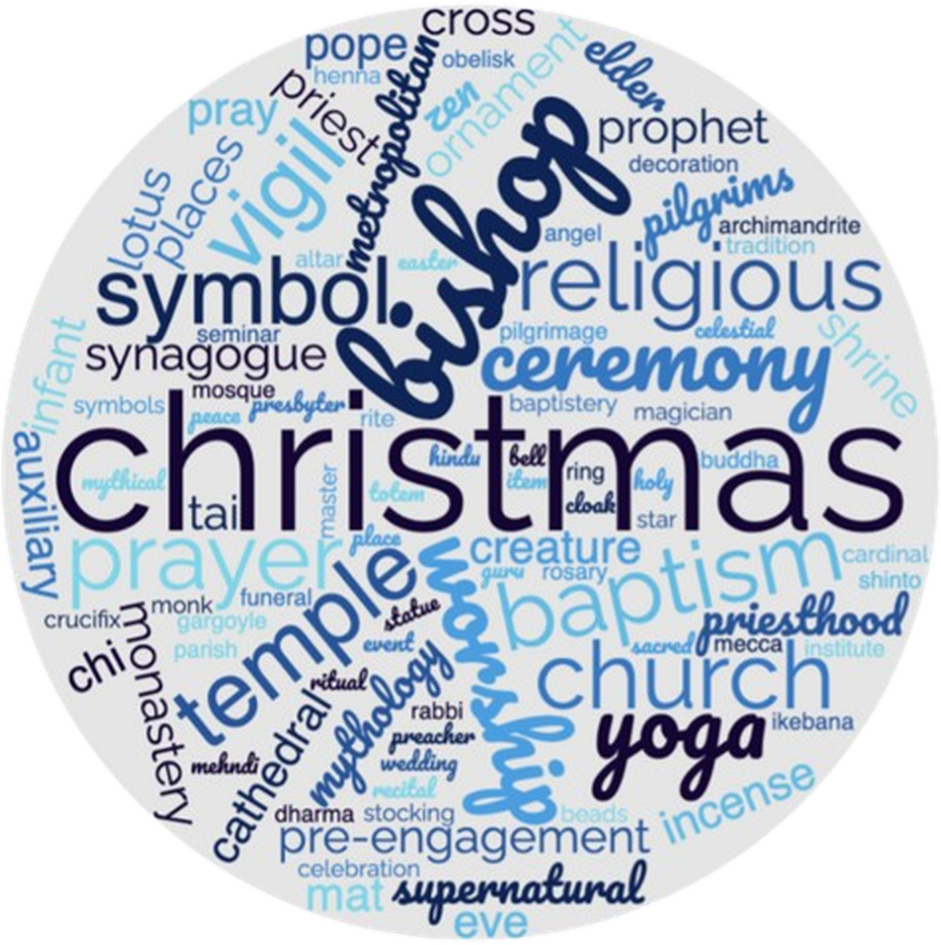


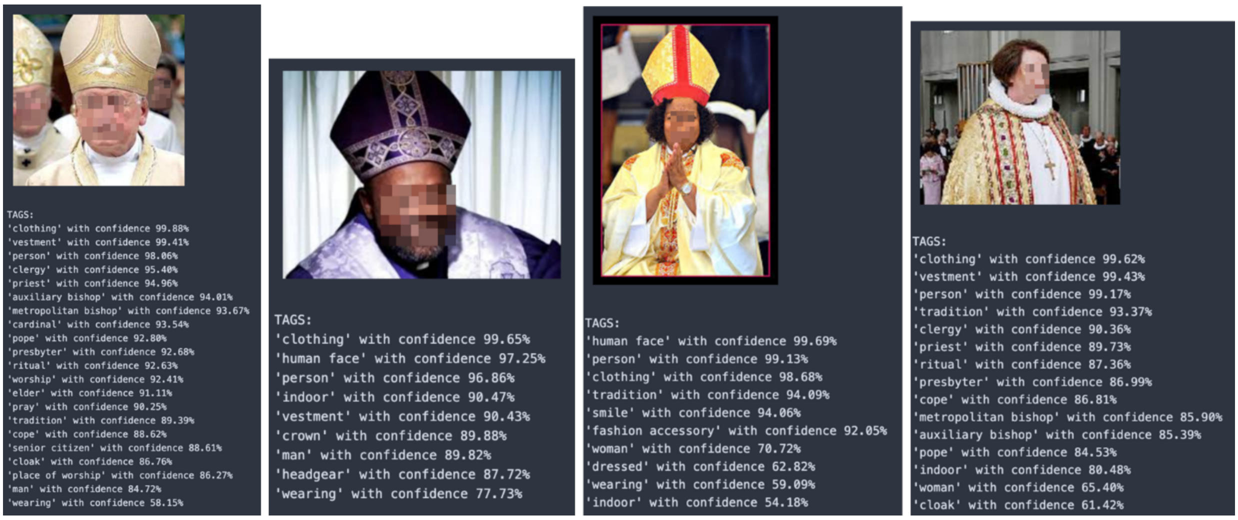
Figure 4. Word cloud by Anton Berg.
Previously published in Berg and Valaskivi (2023a)

The word cloud in Fig. 4 illustrates the overrepresentation of Christian vocabulary. The larger the font in the word cloud, the more frequently the word appears in the material.

A more detailed qualitative analysis of the images and their classification tags shows that so-called ‘high church’ Christianity, in its old European forms, is most readily recognized and receives the most tags.

But there is one very visible exception: If the people who appear in images depicting Christianity are not white, the ability of the systems to recognize the religious context disappears even in a ‘high church’ environment.

Our comparisons demonstrate clearly that it is the skin colour of the people featured in the images that is the key factor: A white female priest receives as many religious tags as a male priest. Images of either a male or a female priest receive no religious tags at all when the priest is Black (see Fig. 5).
Charismatic Christianity is also poorly recognized as can be seen in Fig. 6.



```
TAGS:
'clothing' with confidence 99.88%
'vestment' with confidence 99.41%
'person' with confidence 98.06%
'clergy' with confidence 95.40%
'priest' with confidence 94.96%
'auxiliary bishop' with confidence 94.01%
'metropolitan bishop' with confidence 93.67%
'cardinal' with confidence 93.54%
'pope' with confidence 92.80%
'presbyter' with confidence 92.68%
'ritual' with confidence 92.63%
'worship' with confidence 92.41%
'elder' with confidence 91.11%
'pray' with confidence 90.25%
'tradition' with confidence 89.39%
'cope' with confidence 88.62%
'senior citizen' with confidence 88.61%
'cloak' with confidence 86.76%
'place of worship' with confidence 86.27%
'man' with confidence 84.72%
'wearing' with confidence 58.15%
```

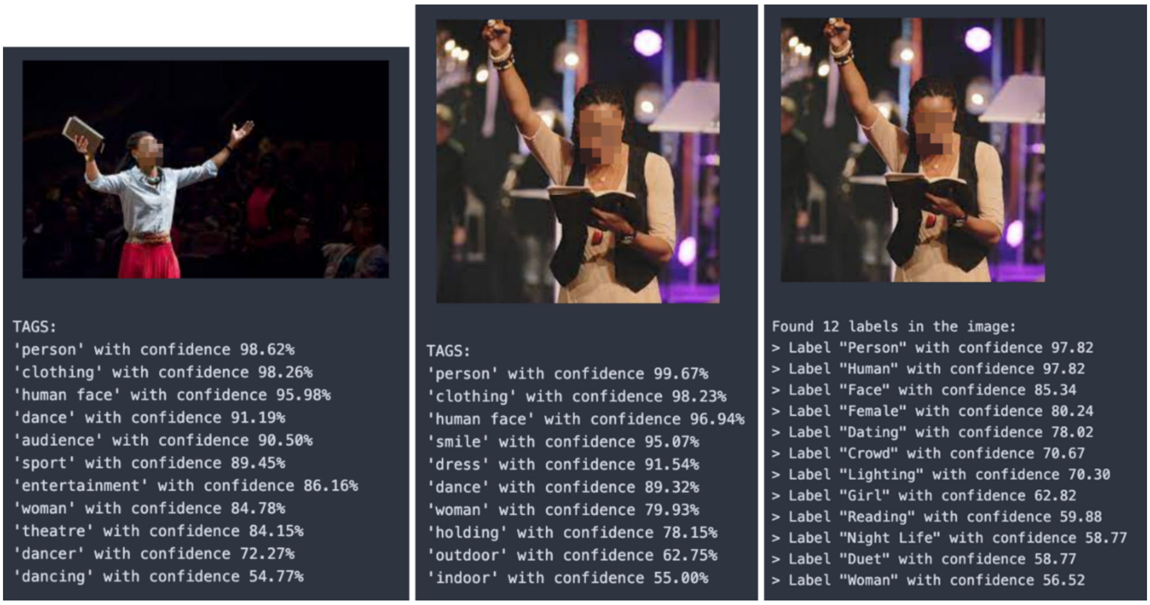
```
TAGS:
'clothing' with confidence 99.65%
'human face' with confidence 97.25%
'person' with confidence 96.86%
'indoor' with confidence 90.47%
'vestment' with confidence 90.43%
'crown' with confidence 89.88%
'man' with confidence 89.82%
'headgear' with confidence 87.72%
'wearing' with confidence 77.73%
```

```
TAGS:
'human face' with confidence 99.69%
'person' with confidence 99.13%
'clothing' with confidence 98.68%
'tradition' with confidence 94.09%
'smile' with confidence 94.06%
'fashion accessory' with confidence 92.05%
'woman' with confidence 70.72%
'dressed' with confidence 62.82%
'wearing' with confidence 59.09%
'indoor' with confidence 54.18%
```

```
TAGS:
'clothing' with confidence 99.62%
'vestment' with confidence 99.43%
'person' with confidence 99.17%
'tradition' with confidence 93.37%
'clergy' with confidence 90.36%
'priest' with confidence 89.73%
'ritual' with confidence 87.36%
'presbyter' with confidence 86.99%
'cope' with confidence 86.81%
'metropolitan bishop' with confidence 85.90%
'auxiliary bishop' with confidence 85.39%
'pope' with confidence 84.53%
'indoor' with confidence 80.48%
'woman' with confidence 65.40%
'cloak' with confidence 61.42%
```

Figure 5. When the members of clergy in the images are white, male or female, the image recognition services produce a rich variety of categorisations related to Christianity. When the clergy members are Black, the systems give no labels related to religion.

The striking categories among the tags in Fig. 6 are are those referring to nightlife, get-togethers, amusement, and entertainment, but also dating and dancing. In this the image recognition services follow the general media imagery trend and present Black women as sexualized objects – already an established finding in Black studies, feminist media studies and racism research (see e.g. Cesaire, 1955; Staples, 1973; Ammons, 1995; Roberts, 1999; Wallace, 1999; Woodard & Mastin, 2005; Hobson, 2005; Richardson-Stowall, 2012; Noble, 2018)



```
TAGS:
'person' with confidence 98.62%
'clothing' with confidence 98.26%
'human face' with confidence 95.98%
'dance' with confidence 91.19%
'audience' with confidence 90.50%
'sport' with confidence 89.45%
'entertainment' with confidence 86.16%
'woman' with confidence 84.78%
'theatre' with confidence 84.15%
'dancer' with confidence 72.27%
'dancing' with confidence 54.77%
```

```
TAGS:
'person' with confidence 99.67%
'clothing' with confidence 98.23%
'human face' with confidence 96.94%
'smile' with confidence 95.07%
'dress' with confidence 91.54%
'dance' with confidence 89.32%
'woman' with confidence 79.93%
'holding' with confidence 78.15%
'outdoor' with confidence 62.75%
'indoor' with confidence 55.00%
```

```
Found 12 labels in the image:
> Label "Person" with confidence 97.82
> Label "Human" with confidence 97.82
> Label "Face" with confidence 85.34
> Label "Female" with confidence 80.24
> Label "Dating" with confidence 78.02
> Label "Crowd" with confidence 70.67
> Label "Lighting" with confidence 70.30
> Label "Girl" with confidence 62.82
> Label "Reading" with confidence 59.88
> Label "Night Life" with confidence 58.77
> Label "Duet" with confidence 58.77
> Label "Woman" with confidence 56.52
```

Figure 6. Image recognition services do poorly with charismatic Christianity. When a preacher is both female and Black, Microsoft Azure, instead of categories related to religion, produces labels that refer to night life, entertainment, dancing, and dating.

So far, we have found that:

1) The studied image recognition services have trouble recognizing religion and have a very limited vocabulary for describing religion.
2) When the services do recognize religion, they produce concepts related to Christianity.

https://doi.org/10.33621/jdsr.v7i154875

3) The services perform best with images of ‘high church’ subjects associated with historical, established Christian traditions and institutions, but fail to identify Charismatic Christianity.
4) The services accurately identify ‘high church’ Christianity only if people featured in the images are white.

In sum, the world’s leading commercial image recognition services ’imagine’ the world as a secular place, and they see religion as Christianity and Christianity as European, high-church and white. A norm of whiteness as well as structural racism are built into these services, which is particularly evident in the case of images featuring religious themes.

Concluded heuristically, these systems reproduce values along the lines of nationalist populist parties in Europe (see e.g. Brubaker, 2017). Like these parties, the services suggest that Christianity is a white religion and the only legitimate religion of Europe. Furthermore, like the nationalist parties, AIs devalue other religions. In contrast to European populist parties, however, the studied image recognition systems are not actively working to expand an identity conflict between Christianity and Islam. Instead, when categorising images with religious content featuring non-white people the systems produce algorithmic racialization (Noble 2018) and representational silence (Hall, 1992) as well as continue the tradition of “symbolic annihilation” (Gerbner, 1972).

5. From content confusion to content chaos?

AI systems are not capable of autonomous thought and self-development but need to be trained, which requires vast amounts of human labour (Suchman, 2007; Ruckenstein, 2023). As they are based on data compiled by humans, AI systems inevitably reproduce and reinforce existing power relations and biases in society. They also simplify complex issues because nuances and human reality do not translate easily into data. The implications of the reversed tool feature is even stronger and more prominent in AIs than in many other systems.

Content-generating AI systems produce text, sound or images using the data available to them mainly by means of probability calculations. They produce what might be described as replica or facsimile content – or “as-if content”. This content resembles a hammer you might buy in a one dollar or 100 yen shop: it looks like the real thing, but falls apart the first time you use it. It is made cheap and looks real, but ends up increasing waste in the world. The currently available text generating AI systems based on large language models are mostly based on the probability calculation principle. This means that they produce the kind of language that will with highest probability seem like real language to a human being with comprehension of the language. In other words, facsimile content may or may not contain accurate information, but the AI systems producing this content are not ‘concerned’ about the relationship between content and reality. In a grim view on the future we might think that if the development of social media brought us content confusion, content-generating AI applications might bring us outright content chaos.

If vague and unreliable noise increases in the communication environment to the extent that everything needs to be doubted, it has disastrous consequences to both everyday living and sustaining trust in society.

6. The value system of innovationism

How do understandings of technology then impact the ways in which it is possible to collectively imagine the world?

Some years ago, I was part of a project that studied ‘innovation journalism’. We interviewed journalists and ‘innovation system specialists’ in the United States, Japan, and Finland. Analysing the interviews, I discovered a belief system or contemporary faith that I came to call ‘innovationism’ (Valaskivi, 2012; Valaskivi, 2021). This faith has four core beliefs:

1) Humans are endlessly inventive and resourceful and can always develop new things and technologies.

2) Because of ingenuity and inventiveness, humans can avert the existential threat presented by climate change, even if takes a last-minute Hollywood solution.
3) It is possible to avert the looming destruction without sacrificing the key values of innovationism: competition, growth, success, and progress.
4) Innovations are a way to resolve any 'wicked problem' without having to question the ideology of growth – and at the same time generate new business, start-ups and profits.

Innovationism has been the driving force behind the development of new AIs and our media environment today, which reproduce old colonial, racist and sexist power structures, as our examples of image recognition above have shown. Some of the problems caused by technologies are unintentional and come about because of a narrow perspective, some because the overriding aim is just to make a quick profit, by any means. However, neither technology nor technology developers are neutral or objective. People working to develop technology can and often do have political views, and code can be written to maintain power hierarchies. Values cut through technology as well.

This is the reason why ChatGPT, a natural language generating app that has attracted much controversy of late, raises much deeper questions than students subcontracting their essays to a machine. These questions include: Why would it be in the interest of universities to give the labour of their staff and students free of charge to help train and further develop often unethically produced for-profit AI systems? Or what will happen to universities' carbon neutrality goals if members of the academic community begin routinely to use these highly energy-intensive AI applications to generate text, sound and images?

The upside among the challenges the new technologies pose is that they not only challenge our imagination but also force to collectively think of the boundaries of humanity, and what is a good society. In the academic context, profound questions revolve around the role of the university in imagining communities, a good society, and future technologies.

7. Living in a void

The flow of news constantly informs us about conflicts and oversimplifications fed by the attention factory: Algorithms are used for purposes of manipulation and quick profits and weaponised for triggering conflicts. Smart devices are eroding people's capacity for attention and concentration and undermining children's learning.

In the attention factory where in principle anyone can have their say, there is paradoxically a severe austerity of attention. In fact, attention is not a thing the reversed tools can measure, since attention is not about clicking or liking or lightning-fast emotional reactions. Attention is about listening and concentrating, about focusing, purposefully looking at and seeing. This is the crux of my argument: that life in the contemporary attention attention economy is in fact life in an attention vacuum. The promise of the internet is that everybody can have their say and voice to be heard. When in everyday social media life, the experience is invisibility to human eye despite all the hearts and comments, figuring out at least some of the reasons for the growing sense resentment in societies saturated by digital media – or anxiety attributed to heavy social media use. After all, without attention and care from others, babies die. Being seen by others is a basic human need.

The attention factory created by human beings is not fit for purpose as a communication environment, either from the point of view of individual well-being or from the point of view of building a sense of community, and it is also destructive to the environment. The good thing is that it is human made, meaning that it can also be reimagined and remade by humans. That will require great care, diligence, and imagination as well as a serious reassessment of our conceptions of technology, humanity, and the environment.

The tasks of increasing understanding and imagining differently, in my view, are among the core responsibilities of the university. This requires an academic community that cherishes its responsibilities to research, think and imagine together and is provided resources, time and independence to do so.

8. From provocation to compassion?

A wise colleague recently pointed out that good questions are more valuable than quick answers. This is why this essay concludes with some – hopefully good – questions:

If the medium is the message and social imagining always takes place through the available media technologies, what kind of communication environment should we imagine and develop to strengthen trust in society and among people(s)? What kind of media technologies and practices would support global solidarity and cooperation and build up democracy and equality? How would we need to rethink technologies to minimize the emissions, pollution and biodiversity loss caused by development, production and sales of digital gadgets and their ubiquitous daily use?

Could we imagine developing religious communities, schools, social and health care services, universities and democratic decision-making without a machine or another to mediate every interaction? Are there areas of life that should not be digitized? And above all: What kind of media and communication environment would encourage all of us, rather than react to provocations, pay more compassionate attention to one another?

Finally: Will there be a point when it is no longer possible to move forward by believing that the next innovation will fix the problems caused by the previous ones? What will happen if we reach that point?

References

Ammons, L. (1995). ‘Mules, Madonnas, Babies, Bath Water, Racial Imagery and Stereotypes: The African-American Woman and the Battered Woman Syndrome’, *Wisconsin Law Review*, 5 (5), pp. 1003–1080.

Anderson, B. (1983). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso.

Berg, A. & Valaskivi, K. (2023a). Dataistunut uskonto: Kaupalliset kuvantunnistuspalvelut uskontoa ”tunnistamassa”. *Teologinen Aikakauskirja*, 128(2), pp. 174–200.

Berg, A. & Valaskivi, K. (2023b). ‘Representational Silence and Racial Biases in Commercial Image Recognition Services in the Context of Religion’, in Lindgren, S. (ed), *Handbook of Critical Studies of Artificial Intelligence*, Edward Elgar, pp. 607–618.

Briant, E. L. (2018a). ‘Evidence on Leave.EU, Eldon Insurance and US Health Insurance Industry, in Supplementary written evidence submitted by Dr Emma L Briant, Senior Lecturer at University of Essex’, Digital, Culture, Media and Sport Committee Inquiry into Fake News: UK Parliament [online]. Available at: https://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/digital-culture-media-and-sport-committee/fake-news/written/84032.html

Briant, E. L. (2018b). ‘Further Supplementary Written Evidence on Cambridge Analytica, Leave.EU and Eldon Insurance’, Digital, Culture, Media and Sport Select Committee Inquiry into Fake News: UK Parliament [online]. Available at: https://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/digital-culture-media-and-sport-committee/disinformation-and-fake-news/written/92380.html

Brubaker, R. (2017). ‘Between Nationalism and Civilizationism: The European Populist Moment in Comparative Perspective’, *Ethnic and Racial Studies*, 40(8), pp. 1191–1226. https://doi-org.libproxy.helsinki.fi/10.1080/01419870.2017.1294700

Cesaire, A. (1955). *Discours sur le Colonialisme*. Paris: Presence Africaine.

CLC (2020). ‘Newly Published Cambridge Analytica Documents Show Unlawful Support for Trump in 2016’, *CLC - Campaign Legal Center* [online]. October 16. Available at: https://campaignlegal.org/update/newly-published-cambridge-analytica-documents-show-unlawful-support-trump-2016

Couldry, N. & Hepp, A. (2016). *The Mediated Construction of Reality*. Cambridge, UK: Polity.

Donovan, J. & Dreyfus, E. (2022). *Meme Wars. The Untold Story of the Online Battles Upending Democracy in America*. London: Bloomsbury.

Einstein, M. (2016). *Black Ops Advertising: Native Ads, Content Marketing, and the Covert World of the Digital Sell*. New York: OR Books.

Gerbner, G. (1972). ‘Violence in Television Drama: Trends and Symbolic Functions’, in Comstock G. C. and Rubinstein E. A. (eds), *Television and Social Behavior Reports and Papers, Volume I: Media Content and Control*, Washington, D.C.: Government Printing Office, pp. 28–187. Available at: https://www.ojp.gov/pdffiles1/Digitization/148976NCJRS.pdf

Hall, S. (1992). ‘Race, Culture, and Communications: Looking Backward and Forward at Cultural Studies’, *Rethinking Marxism*, 5(1), pp. 10–18. https://doi.org/10.1080/08935699208657998

Hall, S. (1997). *Representation: Cultural Representations and Signifying Practices*. London: SAGE Publications.

Hobson, J. (2005). *Venus in the Dark: Blackness and Beauty in Popular Culture*. New York: Routledge.

Kannasto, E., Laaksonen, S.-M., & Knuutila, A. (2023). 'I 🌹🍀🇫🇮 You! : Emojis as Emotional-Political Signifiers in Finnish Election Campaign Discussion Online', in Bui, T. X. (ed), *Proceedings of the 56th Annual Hawaii International Conference on System Sciences (HICSS)*, University of Hawai'i at Mānoa. Proceedings of the Annual Hawaii International Conference on System Sciences, pp. 2370–2379. Available at: https://hdl.handle.net/10125/102925

McLuhan, M. (1964). *Understanding Media: The Extensions of Man*, 1st edn, New York: McGraw Hill.

McLuhan, M. & Fiore, Q. (1967). *The Medium is the Massage*. New York: Bantam.

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

Peterson, M. A. (2003). *Anthropology and Communication: Media and Myth in the New Millenium*. London, UK: Berghahn Books.

Richardson-Stovall, J. (2012). 'Image Slavery and Mass-Media Pollution: Popular Media, Beauty, and the Lives of Black Women', *Berkeley Journal of Sociology*, *56*, pp. 73–100. Available at: http://www.jstor.org/stable/23345262

Roberts, D. (1999). *Killing the Black Body: Race, Reproduction, and the Meaning of Liberty*. New York: Vintage

Ruckenstein, M. (2023). *The Feel of Algorithms,* 1st edn, University of California Press. https://doi.org/10.1525/9780520394568

Staples, R. (1994). *The Black Family: Essays and Studies*. Belmont, CA: Wadsworth Publishing.

Suchman, L. (2007). *Human-machine Reconfigurations: Plans and Situated Actions.* New York: Cambridge University Press.

Valaskivi, K. (2012). 'Dimensions of Innovationism', in Nynäs, P., Lassander, M. & Utriainen, T. (eds), *Post-Secular Society,* New Brunswick, N.J.: Routledge, pp. 129–156.

Valaskivi, K. (2020). 'The Contemporary Faith of Innovationism', in Bell, E., Gog, S., Simionca, A. & Taylor, S. (eds), *Spirituality, Organisation and Neoliberalism: Understanding Lived Experiences,* Edward Elgar, pp. 171–193.

Valaskivi (2022). 'Circulation of Conspiracy Theories in the Attention Factory', *Popular Communication: The International Journal of Media and Culture*, 20(3), pp. 153–161. https://doi.org/10.1080/15405702.2022.2045996

Wallace, M. (1999). *Black Macho and the Myth of the Superwoman*. London: Verso.

Woodard, J. B., & Mastin, T. (2005). 'Black Womanhood: "Essen ce" and its Treatment of Stereotypical Images of Black Women', *Journal of Black Studies*, 36(2), pp. 264–281. Available at: http://www.jstor.org/stable/40034332

JDSR 7(1) (2025) 46-65

JOURNAL OF DIGITAL SOCIAL RESEARCH

10.33621/jdsr.v7i154877

Using Buddhist skillful means

Conducting digital ethnography in diasporic digital Chinese Buddhist communities in Canada[1]

Xiao Han

Université du Québec à Montréal, Canada

🖂 han.xiao@courrier.uqam.ca

Abstract

This study addresses the growing call from scholars, such as Heidi Campbell, for a deeper reflection of methodological approaches to digital ethnography within various religious traditions and communities. In this article, I examine how I utilize a collection of "skillful means" informed by Buddhism, namely a mixed set of digital research methods encompassing reflexive choices and decisions, positioning, and creativities. This set of tools is situationally tailored for and derived from interacting with Chinese Buddhist diasporas in French Canada in the context of digital social media throughout my digital fieldwork. I use ethnographic vignettes to illustrate how these practices, afforded by the Buddhist ideas, digital possibilities, and ethnographic reflexivity, are crucial to constantly navigate, negotiate, and devise new strategies for exploring diverse networked digital field sites through *interconnectivity, fluidity, immediacy* and *disruption* and conducting multi-modal participant observation. By presenting the complexity and intricacy of the insider-outsider conundrum, I highlight key digital features of social media platforms such as WeChat, which can be strategically leveraged by a Buddhist researcher and practitioner to actively shape and present their digital image and voice within the communities they studies. I further reflect on how these dynamics can uniquely influence both the individuals and the communities being researched. Finally, I address the caveats and potential pitfalls this approach could potentially bring about.

Keywords: reflexivity, positionality, Canadian Chinese Buddhism, digital ethnography, WeChat ethnography

1. Introduction

On the morning of August 8, 2021, I opened my WeChat and YouTube and was greeted by an ocean of dazzling messages from various Buddhist communities: on YouTube, the Venerable Master Ru Zhong of Montreal Fo Guang Shan was holding an online filial ceremony[2]; on Telegram, members of the Khenpo Sodargye' Tibetan Buddhism study group were called upon to practice a guru yoga on Zoom. In the WeChat group, the abbot of Cheng Shui Temple thanked the volunteers who offered sugar cane juice to the temple. The guys in the Dharma Art group posted a Bodhisattva emoji that their team just designed

[1] This research has been approved by the ethics review committee involving human subjects (CIEREH) at Université du Québec à Montréal.

[2] A kind of Chinese ritual honouring one's parents and ancestors (influenced by Confucian ideas).

for other group members. The Montreal Prayer Group was making a weekly schedule for the group members who will be on duty next week to host the sutra chanting ritual dedicated to the deceased who died from COVID-19. The Pure Land Sutra Chanting WeChat Group was sharing China's Master Da'an's Dharma Talks video on YouTube about how to chant the name of the Buddha. This is how my everyday online Buddhist life has unfolded since the outset of COVID (fieldnote, 8th August 2021).

During the COVID-19 outbreak in 2020, like many others, I was thrown into self-isolation, an experience that drastically disrupted my usual social connections. This isolation deepened my desire to connect with a community that could offer spiritual solace. As a researcher in the field of Buddhist studies, and at the same time, a female first-generation immigrant and Chinese Buddhist practitioner based in Montreal, I sought to engage with local Han Chinese[3] Buddhist communities to provide spiritual comfort for myself and engage in a community promoting resilience in times of crisis. Concurrently, as a researcher studying Chinese Buddhist diasporas in Canada, I was curious to explore their beliefs, practices, and stories within the digital space. My spiritual and intellectual seeking eventually guided me to establish connection with and subsequently join these six of Han-Chinese Buddhist communities, which thus became the subjects of my research project: 1) WeChat[4] Group of Fo Guang Shan Hua Yan Temple (I.B.P.S. of Montreal) affiliated to Fo Guang Shan, a globally well-known Chinese Mahayana Buddhist organization headquartered in Taiwan; 2) WeChat Group of "The Joy of Chan as Diet", affiliated to Cheng Shui Temple, affiliated to a Montreal Chinese and Vietnamese temple led by Chinese-Vietnamese nuns and attended by mostly ethnic Chinese devotees, catering for the general public who regularly consume vegetarian food; 3) Telegram group of Bodhi Study Society community, founded by Khenpo Sodargye who is affiliated to the Serta Larung Five Science Buddhist Academy in China, one of the largest Tibetan Buddhist academies in in contemporary world; 4) WeChat Montreal Prayer Group, created in memory of a deceased Chinese immigrant BBQ owner, includes reciting Buddhist classics by group members on Zoom on a daily basis for the deceased in Montreal;5)WeChat Group of Pure Land Sutra Chanting community, connected to Pure Land Buddhism tradition in mainland China's Donglin Temple; 6) WeChat Group of "Zen Tea Flavor - Kagyu Center", affiliated to Rigpe Dorje Centre (Montreal), founded in 1987 as the first of many centers to be established by the 3rd Jamgon Kongtrul Rinpoche Lodro Chokyi Senge (Jamg Kongtrul Rinpoche lineage) in North America. It is crucial to underscore that while some of these communities do not only exist online but also have physical entities based in Montreal prior to COVID-19 such as Fo Guang Shan Hua Yan Temple, Cheng Shui Temple and Kagyu Center. The number of members of each group ranges from several dozen to more than 200 individuals. The digital communities are not exclusively reliant on digital platforms such as WeChat, or Telegram, which primarily serve as their community interaction hubs. They also utilize video communication software like Zoom, alongside an array of websites and social media platforms like Facebook, YouTube, and Instagram. These additional channels offer a multiplicity of interaction levels and further avenues for community engagement[5].

[3] Han Chinese refers to the majority ethnic population in Mainland China.

[4] The most popular Chinese messaging app, widely used among diasporic ethnic Chinese, combines features similar to those of Facebook and Messenger, which will be discussed later.

[5] The members of these different communities are primarily first-generation Han Chinese immigrants, mostly from mainland China, with a smaller portion being ethnic Chinese diasporas from other Asian regions and countries such as Vietnam, Hong Kong, and Taiwan. Most of them speak Mandarin, while only a few speak Cantonese. In general, many participants in these groups have obtained Canadian citizenship, and some hold permanent residency in Canada. As the largest international Buddhist organization among the groups I participated in, Fo Guang Shan is highly inclusive and diverse in terms of gender, age, socio-economic status, and educational backgrounds. The predominant participants are ethnic Chinese, with a very small group of local Quebecois. Cheng Shui Temple, known for selling Chinese-Vietnamese food, has attracted mostly mainland Chinese, Chinese Vietnamese, and Cantonese participants at the temple. However, their WeChat group primarily consists of immigrants, international students, and temporary residents from mainland China. At the Kagyu Center, membership is split between Quebecois and mainland Chinese immigrants, but their WeChat group consists solely of Chinese participants, many of whom are younger immigrants in their thirties with stable jobs and decent incomes. Other semi-closed groups, such as the Pure Land Group, Bodhi Study Society, and Montreal Prayer Group, are more homogenous in demographics. They consist mainly of middle-aged, middle- to upper-class immigrant professionals and intellectuals from mainland China, including engineers, professors, and doctors. These members are financially well-off and familiar with up-to-date digital technology.

Their Buddhist practices and understanding of Buddhism also varied due to sectarian differences between Tibetan Mahayana and Han Chinese Buddhist traditions. Except for the WeChat Group of Fo Guang Shan Hua Yan Temple and Kagyu Center, other groups consist of solely lay-led Buddhist practitioner communities, lacking direct monastic involvement in these digital communities. Even though there is no monastic presence in these digital communities, some groups maintain the ability to connect with their Buddhist leaders and teachers in China through digital platforms. Members from these groups have been immersed in Buddhism for many years and have a relatively deep and intricate understanding of nuanced practices, Buddhist philosophies and scriptures[6].

The increasingly entangled relationship between religion and digital media, as illustrated in the opening vignette and the introduction of the six communities, has been extensively researched by scholars over the past decade. Heidi Campbell introduced the essential framework of "digital religion" to understand what it means to be "a religion that is constituted in new ways through digital media and cultures" (Campbell, 2013a: 3-4). Specifically, the profound transformation of religious structures and practices has been deeply reflected in the reality that the global Buddhist world is becoming increasingly digital. Gregory Grieve (2017:6) proposed the concept of "digital Dharma" or "digital Buddhism," referring to "the Buddhism that users encounter on the screen." Drawing on his "Buddhist-informed" ethnographic work on a Zen community in Second Life's digital spaces, Grieve demonstrated how digital media shapes and sometimes challenges conventional understandings of Buddhist identity, community, and practices, and even the authenticity of the Dharma (2017). Alongside Daniel Veidlinger, Grieve (2018) co-edited another key piece of literature-the first volume solely dedicated to Buddhism and digital media, *The Pixel in the Lotus: Buddhism, the Internet, and Digital Media*. This volume explores Buddhist practice and teachings in an increasingly digitalized world. Through various methods, including case studies, ethnographic work, content analysis, and interviews with practitioners and cyber-communities, the contributors examined how contemporary global Buddhism is manifested in digital media. The volume covers digital fields such as virtual worlds, social media, and mobile devices.

The COVID-19 pandemic prompted more research related to online Buddhism and its response to the crisis. Scholars have documented how global Buddhist communities in Canada, the U.S., and Australia responded to the COVID-19 crisis with resilience by transitioning to virtual programs, conveyed through digital platforms such as Zoom, YouTube, and Facebook for sutra chanting and donations (Wilson, 2020; Tseng, 2020; Sang, 2021). While the COVID-19 situation drew significant scholarly attention to global online Buddhist practices and communities, the digitalization of global Buddhism had long been taking place. It is thus critical to recognize that COVID-19 itself did not bring about the burgeoning of digital Buddhism worldwide but rather catalyzed it.

In line with these pioneering research approaches, many researchers working on Chinese-speaking Buddhist communities have added more context and invaluable ethnographic data, making Chinese digital Buddhist communities more visible within the academic horizon. To cite a few, Stefania Travagnin (2019) examined how government involvement in China shapes mainland China's digital Buddhist ritual practices in the temple-developed online platform. She further discussed how a Chinese famous temple used a robot monk to engage with its followers (2020). Weishan Huang (2017) demonstrated how Chinese social media, such as WeChat, facilitates the construction of Buddhist communities and creates a digital sacred space at both global and local levels through the case study of the Tzu Chi Buddhist organization in Shanghai. Yanshuang Zhang(2017) conducted a comparative analysis of how Buddhist and Christian communities in China use Sina Weibo, a major Chinese social media platform, to interact with

[6] Despite the prominent and significant presence of male participants in all six groups, the gender ratio is predominantly female, which strongly suggests that ethnic Chinese women are playing an increasingly substantial role in Buddhist communities in Canada. However, for the sake of focus, I do not intend to explore the gender dimension in this article. It is important to note that my ethnicity, gender, and religious background as a Chinese-origin, female, immigrant Buddhist practitioner have significantly facilitated my interactions with participants. These aspects allow me to communicate freely without the need for an interpreter, build trust and rapport-particularly with female participants-with relative ease, and engage in and observe all relevant activities in greater depth. Further reflections on my positionality and reflexivity will be discussed in detail in the following sections.

participants, building religious communities as well as forming religious identities. Francesca Tarocco (2017) explored how digital technologies, including Weibo[7] and WeChat, influence the dynamics between devotees and Buddhist monastics. These works of literature highlight how Chinese Buddhist communities are influenced and reshaped by advanced technologies and digital platforms.

The COVID-19 pandemic saw a rapid growth in Chinese digital Buddhist practices and transnational communities that are globally networked in a digital world, triggering more related research. For example, Xiao Han's (2022) study focuses on how a Thailand-affiliated Chinese Theravada Buddhist group based in Beijing used WeChat to perform online-offline synchronized meditations to accumulate digital merit in response to COVID-19. Kai Shmushko explores how Tibetan Buddhist communities physically based in Shanghai responded to the COVID-19 pandemic by linking mask-wearing and commercial activities to Buddhist merit, leveraging digital social media and the internet. In more recent work, Shmushko (2021) reviews methodological developments and challenges in the ethnographic study of digital Buddhism in both the PRC and Taiwan, highlighting the significance of including religion, technology, and the market economy in studying Chinese cyber-Buddhism.

Nevertheless, while the aforementioned pioneering research is extremely helpful and instrumental in understanding current digital Chinese Buddhist landscapes, it primarily echoes, and arguably falls into, the essential conceptual framework that Campbell (2013b) pointed out when studying online religious groups-namely, authenticity, community, identity, ritual, and authority. In other words, current scholarship on digital Buddhism mostly focuses on presenting, conceptualizing, and theorizing emerging phenomena and researched subjects, rather than offering down-to-earth ethnographic reflections on how to engage with religious Buddhist individuals encountered on various digital platforms and the impact of the researcher on the digital groups being studied.

More specifically, as I began to explore the study of Chinese sangha on digital platforms, I encountered significant methodological challenges rooted in the everyday practice of digital ethnographic fieldwork, which has rarely been addressed in previous scholarly discussions. These challenges soon evolved into research questions guiding this study: On the one hand, how can I make sound judgments and informed decisions when selecting digital field sites, gaining access, and fostering acceptance? What is the most appropriate approach to conducting participant observation with digital Chinese immigrant Buddhist groups, often organized as membership-based or semi-public networks, with their own distinct discourse, social codes, and preferred forms of digital Buddhist practice? On the other hand, how should I navigate the intricacies of my entry into, engagement with, or even disengagement from digital fieldwork, especially when dealing with the projections, scrutiny, and suspicions from Chinese and Buddhist communities? Moreover, how should I position myself as a lay Buddhist practitioner (insider) and an academic researcher (outsider) as well as a newcomer to their groups?

Unfortunately, although desperately needed, no scholarly attempts have been made so far to engage in methodological reflexivity on how to research diasporic Han-Chinese Buddhist communities in a digital setting, let alone providing comprehensive reflections on how to handle dual positionality as a researcher (outsider) and practitioner (insider) in this context. This methodological predicament has prompted me to respond to the significant emerging appeal for deeper methodological reflections and explorations in digital ethnography to suit the dazzling change in the field. This call—openly made by a group of leading scholars of digital religion, such as Campbell in the *Religion and Digital Media* panel at the AAR conference in San Antonio in November 2023—is a timely response to the rapid growth of AI and algorithms that are increasingly shaping how religious individuals practice and express their faith in digital settings. Therefore, we researchers urgently need to upgrade our methodological tools and reconsider how to navigate these digital religious communities, especially when approaching specific non-Western religious groups that stretch across different cultures, traditions, ethnicities, and geographical regions.

[7] Chinese Weibo (微博) is a popular microblogging platform in China, a Chinese version of Twitter.

https://doi.org/10.33621/jdsr.v7i154877

To address this methodological gap, I draw on a key Mahayana Buddhist concept, *Upaya* (方便, *fangbian*), meaning "expedients," "stratagem," or "skillful means" (hereafter), which is highly emphasized in one of the key Mahayana Pure Land Buddhist scriptures, the *Lotus Sutra*[8] (Williams, 2008). This concept refers to the Buddha's pedagogical flexibility and wisdom in adapting the teachings to suit changing circumstances when teaching the Dharma to various recipients from diverse geographical, cultural, and linguistic backgrounds, using different similes, parables, or referencing the audience's rituals and traditions, ultimately leading them to understand the Buddhist truth (Keown, 2005:18). Beyond its pedagogical applications, the concept of skillful means is also interpreted from an ethical perspective to encompass any behavior performed by the Buddha, Bodhisattvas, or even all Buddhist practitioners out of compassion, wisdom, and a willingness to benefit others (Williams, 2008:15; Keown, 2005:18). Furthermore, from the perspective of daily practices and moral evaluation, skillful means encourages Buddhist practitioners to act in accordance with the spirit of the Dharma rather than adhering to fixed, predetermined principles and precepts when it comes to lived Buddhist praxis (Keown, 2005:18; Schroeder, 2004:150).

I also draw on the key anthropological method of reflexivity. Since the 1980s, scholars in Religious Studies have begun to integrate the principles of reflexive anthropology, a paradigm that challenges the traditional notion that anthropologists can create objective knowledge about the participants they research. In *Writing Culture*, a seminal work by Marcus and Clifford (1986), the authors acknowledged the influence of anthropologists' own backgrounds, perspectives, and stances on the knowledge they construct, illustrating that reflexivity is essential and integral to the role of anthropologists. This incorporation of reflexivity naturally led to an examination of the boundaries of truth claims. I thus argue that ethnographic reflexivity is closely associated with skillful means when examining Buddhist communities, as it deeply embodies adaptability, ethical sensitivity, and recognition of the diversity of participants, and it naturally resonates with the core ideas of skillful means.

Therefore, aside from its soteriological aim of tailoring teachings to help sentient beings attain enlightenment, I regard skillful means as a Buddhist-informed model guiding my entire fieldwork-a model comprised of a collection of ethnographic practices, shaped by a set of reflexive choices and decisions. The guiding spirit of this model, much like skillful means, is driven by Buddhist insights such as compassion and a Buddhist-informed sense of responsibility. This approach is rooted in my dual role as both a Buddhist practitioner and an academic. As a practitioner, I aim to practice Buddhism with a soteriological pursuit, while using my academic work to benefit Buddhist communities by increasing their visibility in the Western academic world. As an academic, my fieldwork is grounded in a dynamic, ever-changing digital field, which demands a highly adaptable, strategic approach, along with exceptional ethnographic sensitivity, reflexivity, and creativity. In light of my dual roles, the advantage of this model is that it has led me to engage with communities that embody the very teachings of my guiding model in their daily lives.

As a result, rather than attempting to create a new set of methods for digital ethnography, this article aims to reflect on the situational, tailored, and constantly responsive methodological decisions, choices, and inspirations associated with ethnographic conundrums when conducting digital ethnography with Chinese diasporic Buddhists in Western countries such as Canada. Specifically, I aim to investigate how the researcher's role and voice are shaped by cultural, social, and religious nuances and distinctions when studying Chinese Buddhist immigrant communities through digital ethnographic fieldwork. I further aim to demonstrate that conducting digital ethnographic work with Chinese Buddhist communities requires not only leveraging reflexivity and positionality to effectively engage in fieldwork within a digital,

[8] The Lotus Sutra is one of the key Mahayana Buddhist texts. It is well-known for its inclusive teachings on the universality of Buddha-nature, the potential for all sentient beings to attain Buddhahood, as well as its emphasis on skillful means in teaching different audiences to lead them to the ultimate truth.

diasporic context, but also an understanding of how the social and cultural perspectives and standings of both the participants and the researcher can intersect and shape the fieldwork process.

Furthermore, it is critical to reflect on how these dynamics can specifically influence the researched individuals and communities in a distinctive manner unique to the digital setting. Following these reflections, I argue that my positionality in the digital field extends beyond merely being a fellow Buddhist participant among community members; instead, I have actively established my researcher's visibility and voice using digital affordances, positioning, and creativities that are situationally tailored for and derived from interacting with Chinese Buddhist diasporas in Canada in the context of the digital realm throughout my digital fieldwork.

This paper draws on skillful means, along with anthropological reflexivity, as the guiding ethnographic model for my ongoing digital fieldwork on six communities from the beginning of COVID-19 in 2020 until early 2023. Employing brief ethnographic vignettes, the following will be a mixed presentation of ethnographic reflections and fieldwork data. I will first explore how I practice ethnographic reflexivity and skillful means to identify six digital field sites and how specific digital participant observation was conducted through alternating modes. I will then explore how I skillfully navigate beyond the dichotomous roles of researcher (outsider) and lay practitioner (insider) in the digital Buddhist communities. Moreover, I will delve into how I leverage my "researcher voice," or the visibility of my academic expertise within the groups (a term to be developed and explained later), to engage with and impact the communities I study. Lastly, I will discuss how my status as a Buddhist academic projects misconceptions and presents caveats during digital fieldwork.

2. Following the flows of digital Buddhist communities - Navigating the fields

The digital field significantly requires skillful means to navigate effectively and seamlessly. Regarding how to define the digital field itself, I draw on Hine's (2015) arguments on using reflexivity in choosing field sites when conducting digital ethnography. She advocates for an ethnographer's ability to exercise discerning judgment in selecting field sites within the mutable and interconnected digital environment. It is vital that this agency is reflected in the narratives the ethnographer constructs regarding the participants and the field. Additionally, Hine emphasizes the importance of the ethnographer's reflexivity in evaluating how their involvement shapes the field, as well as the impact of their own subjectivity on the relationships with those under study. With this in mind, throughout my research, I have actively adopted reflexivity in each interaction within my research communities, from the choices of the digital field to the modalities of participant observation, as discussed in what follows.

2.1 Where is the field?

The first aspect of this skillful-means-guided reflexive journey pertains to my identification of digital field sites, which was driven by a dynamic and purposeful choice. By dynamic, I mean that the selection was adaptable and changed as required, and by purposive, I mean that it was intentional and based on prescribed objectives in terms of the nature of groups such as their ethnicity (ethnic Han Chinese), sectarian practices (Tibetan Nyingma Buddhism, Humanistic Buddhism, Pure Land Buddhism in Chinese Mahayana traditions) and their geographic locations (physically based in Montreal) as well as their predominant digital presence. In general, my choice of field sites is guided by the principle that ethnography should be purposive rather than passive, as argued by Hine (2015). According to Hine, ethnographers should not simply follow what the field dictates or stick to predetermined subjects. Instead, we should recognize that "the shape of the field is the upshot rather than the starting point and is the product of an active ethnographer strategically engaging with the field, rather than a passive mapping of a pre-existing territory or cultural unit" (Hine, 2015: 54).

https://doi.org/10.33621/jdsr.v7i154877

Guided by this principle, in my fieldwork, I identified four factors associated with the nature of the digital world that shaped my group choices in the field, which I term: *interconnectivity, fluidity, immediacy*, and *disruption*. I will briefly introduce and illustrate these terms with ethnographic data as follows. *Interconnectivity* means different communities across various digital platforms can be fluidly interconnected, which "mirrors the experiences of navigating through a connected world" (Hine, 2015: 122). Buddhist practitioners belong to different Buddhist digital groups, which can serve as linking nodes to invite each other to a new group - it is not surprising to find that the same group of members in one group also dwell in another group. Therefore, by harnessing the connectivity of people on social media, I serendipitously and gradually discovered other, more relevant field sites that warranted more in-depth research. For example, I discovered and then became a participant in the Pure Land Chanting Group simply because one participant from that group posted an announcement in the Cheng Shui Temple WeChat Group, looking for individuals interested in chanting Pure Land sutras together on Zoom. Following this, I added him to join the WeChat group. *Fluidity* indicates that the digital Buddhist communities I researched are highly fluid and constantly in the making, with individuals moving in and out, or becoming completely dormant after a period of active interaction, and then suddenly returning to bustling engagement again, especially in some lay-led communities with loose regulations, such as the Montreal Prayer Group. For example, after a peak of prayer rituals dedicated to COVID deaths, as the death rate significantly decreased, the prayer group became so quiet that it seemed to be a "dead group." During its dormant phase, I simply stopped participating, as it did not generate data at that time.

Immediacy illustrates situations where a significant event demands immediate attention, prompting an investigation into an unfamiliar digital platform that typically falls outside the scope of my regular exploration, or even the creation of new groups merging familiar participants from various communities, like water flooding in. For example, I saw someone post a WeChat group QR code and promote photos of an upcoming group's offline Meditation Tea event at the Kagyu Center in multiple Chinese immigrant WeChat groups on Chinese New Year's Eve 2023. After I joined the group, I found that the number of members in this newly created WeChat group increased dramatically from a mere dozen to over a hundred in a week and it subsequently became one of my regular field sites. Finally, the *disruption* or unexpectedness of the digital field sometimes caught me off guard, leaving me with no choices—the field suddenly disappeared overnight without any warning or explanation, which is usually hard to imagine in an offline context. Sometimes the technical affordances of the digital environment create unique "surprises" in the online world. This was reflected in my personal experience when a subgroup of the Bodhi Study Society on Telegram was removed and disbanded by the organizers in two hours, as a preemptive strategy after a digital scam emergency (presumably due to political infiltration). Similarly, I also experienced the behaviour of a gatekeeper who initially displayed considerable kindness and enthusiasm towards me, but later mysteriously changed her demeanour and removed me from the WeChat group where I intended to conduct my field research.

2.2 No longer familiar strangers anymore - Online-offline connection

This is the second aspect concerning navigating the digital field. Much like navigating digital field sites demands adaptability to a constantly shifting environment, the digital interactions likewise increasingly blur the once-clear boundaries between online and offline religious experiences and thus mandate a mindset that avoids viewing the online and offline with fixed boundaries. Hine (2015) highlighted the importance of studying the Internet experience that is integrated into people's everyday lives, indicating that the researcher may engage with the field through various means, including mediated interactions online or face-to-face engagements offline, or a combination of both. Heidi Campbell (2012) argued that one cannot ignore the offline aspects when exploring online communities because, in fact, there are no fixed boundaries between the two realms, and they are occasionally convergent at some points. With these insights in mind, I emphasize that offline engagement is pivotal in achieving a multidimensional or multi-layer understanding of online researched individuals or communities for the following reasons.

First, the transition from online to offline is inevitable. Although most of my fieldwork was done in the digital space, the communities I am looking at have a physical base or at least had one before in Montreal. As lockdown restrictions were lifted and people started to engage in offline gatherings, there was a transition from digital to in-person events. These included welcoming new members, vegetarian food sales, Mid-Autumn Day Mountain pilgrimages, and regular practices such as communal "Eight Precepts Retreats" and Buddhist etiquette courses. These events naturally yielded significant ethnographic data through offline participant observation. Second, by being seen as a real participant person by others and being physically present, I further validate my membership and build trust in the group. Third, offline engagement significantly bolstered my rapport with certain members of the community who I had previously known only as "familiar strangers" through online interactions. Some of the online gatherings within groups such as Bodhi Study Group do not involve any cameras but only voice communications due to the fear of political persecution associated with participation in Tibetan Buddhism, as its global Buddhist leader such as Dalai Lama in China is portrayed by CCP as being involved in separatism[9]. Being able to associate faces with the voices and build new connections with previously unknown participants proved to be immensely beneficial. Moreover, it enables a more introspective approach, allowing me to reflect on how individuals in the network portray their Buddhist identities online and to explore the tensions this may create with their offline identities (Bluteau, 2021). Last but not least, the bonus of offline participant observation often featured unexpected, interesting incidents and nuanced stories, which enriches my analysis and interpretations of the Chinese Buddhist digital sangha.

2.3 Multi-modal digital participant observation

A third characteristic of the digital field entails responsive and situational multi-dimensional participation. Throughout my online and offline fieldwork, from late 2020 to March 2023, I actively participated in a diverse range of activities within the researched groups. These activities included, but were not limited to weekly online Dharma services, chanting, Buddhist ceremonies, Buddhist lectures, conferences, daily practices like visualizations and meditations, and weekly seminars as a regular member. However, given the predominant online nature of my fieldwork, it is important to highlight the everyday participant observation modes that characterized my digital research experience. The modes of digital participation I employed included active participation, engaging with one another, using various mediums such as online text, audio, and video, and interactive social media features to enhance visibility. Additionally, they encompass more subtle forms of presence like dwelling, as well as less noticeable strategies like "lurking" (will be defined later), or salient participation.

In my fieldwork, active participation did not only entail investing significant time and energy in attending important events, it also involved actively engaging with specific informants who were eager to discuss and share their Buddhist experiences online. This required consistent and attentive engagement on a daily basis. This included regularly following their online activities, liking their posts, commenting, or forwarding them, in order to create a sense of "being there" and maintain a continuous presence, ideally in a prompt way to capture the potentially fleeting attention of participants by promptly responding to their posts, signaling my interest and engagement with their content. Dwelling serves as a means to

[9] Buddhism in cyberspace plays an important role in what Chinese-French Buddhist scholar Zhe Ji referred to as a "social force" in China in terms of social mobilization (Kai, 2023; Ji, 2012), which is deemed a substantial threat to the CCP leadership, despite "its perceived docility and its lack of association with foreign imperialism" (Poceski, 2016: 91). This concern is particularly acute when "Buddhism obviously became a means of protest against the rule of the Chinese state," especially when "Tibetan nationalism was loudly pronounced by the political activities of Tibetan monks and nuns" (Yu, 2013, p. 4). The founder of the Bodhi Study Group, Khenpo Sodargye, suddenly disbanded the branches of the Global Bodhi Study Society on December 31, 2019. It later went underground, renamed and rebranded as a novel community in which I participated. Most participants believed that Sodargye, as a Tibetan monk, must have faced considerable political pressure from the government due to his increasingly expansive religious influence among mainland and international Chinese.

https://doi.org/10.33621/jdsr.v7i154877

establish a subtle presence in the digital world, facilitating identity establishment and conducting interviews (Boellstorff et al., 2012: 76). While not obligatory, it can signal long-term commitment and generate favor within the community (*ibid*.). For me, dwelling represents a minimum level of occasional participation. By appearing during major events and engaging in activities, like expressing condolences or sharing greetings on Chinese New Year's Eve, I let community members know of my presence and lay the foundation for further interaction.

In my research, I treated lurking as an alternative mode of silent participation. Lurking, as highlighted by De Seta (2020: 85), is a customary and widely accepted approach to engaging with the digital world, applicable to both researchers and everyday users. Lurking is seen as "a possibility alongside practices such as ignoring, reading, liking, commenting, sharing, editing, and linking" (De Seta, 2020: 86). I use lurking mode for two main reasons. Firstly, in certain Buddhist communities where restraint and mindful speech are encouraged, lurking becomes particularly significant and dominant. It aligns with the expectation for individuals to always be mindful of their body, speech, and mind, according to the Buddhist idea, of following the community's norms by minimizing engagement. Secondly, in larger communities with more than 50 members, the general code for newcomers is to silently participate before actively engaging. This allows individuals to understand and learn "the community's social codes" (De Seta, 2020), internalizing social norms over time, which is also a form of participation. Meanwhile, dwelling typically occurs in smaller groups where a relationship has already been established or where community members are aware of my presence through occasional participation or the presence of my profile avatar. To embody a subtle presence, I carefully chose and crafted my profile avatar, which serves as a manifestation of my identity, passion, inner world, humour, irony, hobby, and social life.

The modes of participation in my research are not mutually exclusive and can adapt as situations evolve. When joining tight-knit communities, I initially observe the atmosphere and follow established decorum. Upon joining the group, introducing myself as a "Buddhist practitioner" and "researcher" serves as a rite of passage of my acceptance in the group, often accompanied by warm welcomes and pleasantries, exemplifying active participation. However, as time progresses, I strike a balance to avoid becoming overly engrossed. I strategically transition into dwelling or lurking modes, maintaining a presence while allowing for measured observation and reflection. These different modes of participation are not entirely contrived as research methods; rather, I argue, part of them are spontaneous responses to the dynamics of the digital community, mirroring the practices of other community members.

2.4 Gradual disengagement

Saying goodbye to fellow digital participants is the most challenging part, demanding an exceptional degree of care and empathy. While disengagement from the field is crucial for further critical and analytical reflection, I regard it as an ethical imperative for a researcher, especially one who is also a Buddhist insider, to assist informants in processing this disengagement. A proper disengagement, with reflexivity and sensitivity, is associated with the "careful consideration of responsibilities and obligations" to the research subjects (Labaree, 2002: 115). The sudden intensive appearance in people's lives for a few years, followed by an abrupt disappearance, can make people feel exploited or betrayed (Zayed, 2021). This is also due to the fact that some of my informants became real friends over time, making it even harder to abruptly sever ties with them.

To address this, I skillfully leverage digital settings by adopting a more gradual approach to disengagement. I maintain loose connections with the field and selectively participate in events that require the presence of all community members (e.g., a New Year gala on Zoom). Furthermore, I use my digital visibility by sharing aspects of my daily life, such as zoo visits or ceramic painting workshops, giving the community a sense that I am still present. This practice of maintaining visibility, which I previously discussed as labour (in the positionality section), began to feel less like a task and more akin to forming bonds of kinship (Abidin, 2020: 67).

3. Beyond the "insider illusion" in digital diasporic Chinese Buddhist communities

Navigating the boundary between being an insider and outsider within a Buddhist community is another excellent example of how skillful means can be applied. Traditionally, anthropologists have employed the study of boundaries as a valuable tool "for discovering who is and what it takes to be accepted as an insider, and to see how, and how strictly, these boundaries are formed and maintained" (Bowie, 2019: 114). However, the incorporation of reflexivity in Religious Studies raised issues such as "the insider/outsider dichotomy [that] does not work precisely because there are no stable categories" (Katie et al., 2015), given that individuals within religious communities are in a constant state of flux across boundaries, and researchers themselves might be practitioners of a particular religious tradition, which grants them a degree of insider status. Additionally, conducting research on religious individuals may lead to instances where participants contest the researcher's interpretations of their practices and beliefs, which "raises questions about representation, and power" (*ibid*.), thereby adding a layer of complexity to the demarcation between insiders and outsiders.

Aside from charting digital field sites, I also realized that ethnographic positionality presents another significant challenge in digital ethnographic work, where the dichotomous distinction between "outsider" and "insider" within a particular digital religious community can sometimes blur or even become irrelevant, requiring more delicate positioning. This is primarily because, under certain circumstances, maintaining an outsider's perspective can become unfeasible. This mirrors Christine Hine's assertion (2015: 85) that "ethnographic research carried out in and of and through mediated communications is always to some extent 'insider research,' since the ethnographer is employing the very means of communication that are simultaneously the object of study."

In my study, this is particularly true because almost every digital group — though exceptions exist — or communities of Chinese Buddhists on social media platforms has certain forms of gatekeepers. It is virtually impossible for someone to gain entry into these digital communities without a certain degree of familiarity with or belief in Buddhism or Buddhist friends or family members. Without this prior knowledge or faith, individuals are often viewed with suspicion, perceived as engaging in religious voyeurism, and their motivations for wanting to join are questioned, if they somehow manage to join one Chinese Buddhist group. Moreover, the political sensitivities surrounding religion in China mentioned previously, especially in the context of Tibetan Buddhism, coupled with a pervasive sense of insecurity stemming from fears of being targeted by China's government due to involvement with Tibetan Buddhism and controversial Tibetan Buddhist figures such as Khenpo Sodargye, only serve to fortify the barriers constructed by these gatekeepers. Thus, in this context, some degree of "insider" positioning is practically a prerequisite for conducting this type of research.

However, I also recognized the "insider illusion", namely I automatically assume the role of a well-accepted insider to a group due to frequent pleasant interactions with community members on social media, does not always work and can suddenly shift in more nuanced contexts. This realization echoes Aston Katie's contention that "the dichotomy of insider/outsider presumes a fixed personhood, an unrealistic assumption that does not account for personal growth or situated experience" (Katie, 2015: 10). Under these circumstances, the implications of an "outsider" could be multifaceted and multi-layered. While my status of being a Buddhist practitioner, a Chinese immigrant, and a Mandarin speaker may generally mean an insider to them, but on a deeper level, my access, belonging and even my pedigree were being scrutinized and seen as an "outsider", and hence my ability to engage in further exchange with these participants was limited. For example, in the Bodhi Study Society, the community primarily centers on the Nyingma tradition, with most members focused on studying Nyingma scriptures and practices. However, there are smaller subgroups and thus relatively marginal groups within this community, such as the Pure Land Buddhism practitioners, and had somewhat less desirable relationships with other subgroups. These underlying tensions often shape the inner boundaries of defining an insider.

https://doi.org/10.33621/jdsr.v7i154877

In a Telegram message exchange on 20 March 2023, I was taken aback when a community member, who had previously praised me as a promising young Buddhist talent and hosted the 2021 Zoom Christmas gala where I performed a Sanskrit Buddhist song, decisively declined my invitation for a Zoom interview without even any hesitation. Given his past enthusiastic support and appreciation for me, I never expected such a response. Later, I learnt that he was a practitioner from the Pure Land Buddhism Telegram subgroup. Then his refusal made sense to me as since I was in the main subgroup primarily studying Nyingma scriptures, he likely considered me as an outsider with respect to my subgroup belonging. Another example is associated with the study of Ke Cui (2015), who shed light on how a fieldworker's pre-existing relationship with some of the interviewees might change due to shifts in insider/outsider positionality during interviews, especially within the context of China's social value system. In my case, I recognized that when "face" (an act of doing a favour) is given to a Buddhist fellow practitioner through the acceptance of interviews, it does not necessarily mean that everything following would go smoothly and be shared in a friendship or an insider setting, as the dynamics are different in a researcher-researched relationship.

For example, I encountered a situation during a Zoom interview where I was on the verge of being relegated to an outsider. This occurred when the interviewee, who had initially introduced me warmly to the Pure Land Sutra Chanting group and often had online exchanges with me in the group, unexpectedly requested that I explicitly disclose my affiliated Buddhist lineage at the start of the interview to decide whether to proceed with the interview. To secure my insider position for the interview, I explained my Buddhist journey since the age of 18, which was not something I typically do in an interview. Another example is, in an unexpected informal conversation, I encountered a middle-aged male Buddhist practitioner who had only recently started his Buddhist journey for two years. He abruptly labelled me as an outsider to Buddhism upon knowing my researcher identity, claiming that if he were presented with my writings, he would not even "bother to read them" because he arbitrarily presumed that I only engage in *studies* of Buddhism rather than *practicing* Buddhism and that my supervisors are all Western scholars who know nothing about Buddhism, which renders me as a "fake Buddhist".

My fieldwork experience gradually taught me that I should not arbitrarily assume myself to be either an insider or an outsider because the distinction between "insider" and "outsider" becomes blurred from time to time. This insight is reflected in the work of Kim Knott (2010), who emphasizes the fluid and shifting nature of the insider-outsider dichotomy in religious studies. She urges the recognition of the crucial role of reflexivity, negotiation, and mutual understanding in the interplay between researchers and the communities they investigate. Knott writes: "My own view, formed in the context of developing a spatial methodology for the study of religion, is that all interlocutors – whether secular observers, religious participants, or those who strategically move between the two positions – are actors within a single knowledge-power field (Knott 2005). Despite their differing goals and interests, they have together defined, constituted and criticized 'religion' in general, particular 'religions' and their beliefs and practices, and the secular or non-religious domain beyond religion." (Knott, 2010, p. 270). On the other hand, I also came to realize that, in terms of a dual role as researcher and a Buddhist practitioner in religious communities, I "cannot escape being both insiders and outsiders" as Wilkinson and Kitzinger acutely observed (Wilkinson & Kitzinger, 2013: 252), because I am on an equal footing, as an actor and participant, alongside other community members, the only distinction being my ethnographic insights. The study of Chuan Yu (2020) on Chinese online translator communities further exemplifies the fluidity of positions. Yu observed that the way a digital ethnographer and her informants position themselves relative to each other is highly contextual and unpredictable, owing to the fluid and ad hoc nature of online practices and communities. In addition, the relativism of the insider-outsider spectrum should also be taken into consideration, as anthropologist Fiona Bowie (2019: 125) points out, "insider and outsider are relative terms", and according to Robert K. Merton (1972: 22), each individual possesses "not a single status, but a status set".

How then should I navigate my position in fieldwork and research? Chuan Yu's (2020) concept of "multiplex persona" resonated deeply with me. It offers a perspective that "views positionality as a decentered entity that encompasses our multi-faceted characters, roles and aspects of identities, presented to and perceived by others and ourselves in the momentary communicative events" in digital space. With this insight, I introspectively examined my own kaleidoscope of positionalities. My manifold social identities that were clearly declared, as I introduced in the introduction when making my first entry into the community, include facets such as a first-generation Chinese immigrant, a doctoral student in religious studies as well as a Buddhist practitioner. Furthermore, my multifaceted engagement with others unveils a diversity of roles, including, but not limited to, being a community member, researcher, ethnographer, event manager, authority or apprentice in Buddhism, consultant, gatekeeper, listener, empathizer, confidante, volunteer, guest presenter, a wife, someone perceived as fortunate, and a young professional. These personas fluidly intertwine, sometimes simultaneously, sometimes separately, as community members engage with the varied dimensions of my identity. However, reflecting on my role as a Buddhist practitioner does not mean limiting my social persona to one aspect but it means being very aware of some blind spots.

It is crucial to recognize that declaring multiplex personas or positionality is not a once-for-all permanent solution for a representation of myself. My understanding of reflexivity in my research further demands me to delve deeper, seeking reciprocal understanding, and reflecting on how misunderstandings are either resolved or contribute to the subjective interpretations and mutual projections in my research. It also involves the discernment, orchestration, and negotiation of these facets of identity during every interaction and decision-making process within the dynamic landscape of the digital realm as can be seen in the next section. This becomes even more salient when engaging with Chinese individuals, for whom *guanxi* ('interpersonal relationships') and *mianzi* ('face", or 'social standing') are deeply embedded cultural values. To gain acceptance and build rapport with Chinese Buddhist communities, especially in digital ethnography, demands that I deftly navigate these sociocultural currents with a measure of tactfulness and sensitivity.

4. Presenting myself as a digital Buddhist researcher rather than a Frenzy Devotee

Crafting a digital identity as an academic Buddhist researcher is itself a form of skillful means. An ethnographic researcher must, as Christine Hine warned: "pay considerable attention to their self-presentation. Establishing one's presence as a bona fide researcher and trustworthy recipient of confidences is not automatic" (2015, p. 20). Gaining acceptance tends to be perhaps sometimes even more difficult in the digital world than offline world, "where a panoply of methods for communication can be used to ingratiate oneself into a community" (Bluteau, 2021: 238). Understandably, some people tend to see or interact with the real person before they build trust and relationships. This requires extra effort and the exercise of caution. In this sense, personal presentation is more essential in the digital world than in the offline world (Bluteau, 2021; Horst, 2009). Throughout my fieldwork, I inadvertently employed a strategy that was described using Crystal Abidin's concept as "visibility labour" — a flexible strategy "enacted to flexibly demonstrate self-conspicuousness" (Abidin, 2016: 90) "in order to win favour among your audience" (Abidin, 2020: 62) through both "physical interaction" and "digital traces" (Abidin, 2020: 60).

This approach was essential in familiarizing my community members and informants to my digital presence and identity, and in establishing a sense of credibility and trust from the outset, which is particularly important in the digital milieu. Because social media such as WeChat, Telegram, and Facebook lack the traditional physical embodiment, in such a setting, visibility entails leaving digital footprints for a potential audience and establishing a trustworthy virtual presence. They were able to evaluate or verify my academic creadibility and social roles, learn more about me through my posts on daily life, or even occasionally discover mutual acquaintances through likes and comments, thus increasing my credibility. My work in visibility labour made it easier for other participants to understand my academic life and research interests, thus distinguishing me as a scholar rather than an apologetic or overly zealous Buddhist devotee. I also regard this as a constant ethical declaration of my research agenda with the community. Furthermore, one of the benefits of showing my status and posts on WeChat to my informants was that this involves deciphering each other's language and skills, evaluating each other's social contexts, balancing statuses, and understanding the spaces that separate us (Abidin, 2020). The aim was to mutually benefit from the social capital and foster "relational care", as Abidin noted (2020: 73).

Besides, a certain degree of visibility allows me to create a "cohabitation" status (Bluteau, 2021) with the community members where I can experience what they are experiencing. Joshua Bluteau further (2021: 268) acutely pointed out that "developing a digital self as a tool through which to access and research the digital field site is powerful. The beauty of this method lies in its dual function as both an access point and a research tool, but by engaging in the same activities as one's informants, a degree of reflexivity can be brought to bear. Furthermore, it is "possible to gain an understanding of the habitus of one's informants and even to cultivate a shared understanding of said habitus if the process of crafting the digital self is sufficiently immersive over a long enough period of time" (Bluteau, 2021: 272). I argue that this "visibility labour" is critically important in building rapport with a small Buddhist group characterized by a greater degree of personal intimacy and transparency who are setting clear boundaries for outsiders and insiders.

The intersection of the "field" with my private life presented a delicate balancing act which entails substantial "behind-the-scenes labor" (Abidin, 2020) - because social media, such as WeChat, is both my fieldwork sites when it comes to the Buddhist groups researched, and my personal communication channel with friends and family. This labour involved making critical decisions regarding how to craft and present my visibility because this visibility is a double-edged sword. While it could foster a sense of relatability and connection, it was vital to exercise caution regarding the non-Buddhism-related aspects of my life that I chose to share with the Buddhist groups and their members. I was mindful that my posts have the potential to elicit a range of reactions from my informants – from resonance and intimacy to skepticism and hindrance. This, in turn, could have potential ramifications on my fieldwork or interviews. Furthermore, scholars have found that the personal and professional entanglement via social media and its intrusion into researcher's personal lives has become a prominent challenge in the realm of social media research (Zayed,2021: 56; Dodds, 2019: 733; Käihkö, 2020: 85). This was also true in my case. For example, the visibility and easy accessibility of social media (during and post fieldwork) sometimes became tricky when I needed to distance myself from the community. Various community members could easily reach out to me through direct messages or video calls, inviting me to unintended socialization, group activities, or volunteering opportunities, often interrupting my personal time. They knew I would check WeChat, and it felt unethical to pretend I wasn't. This became even more frustrating when I was occupied with conferences, thesis chapters, or home-calling my family through WeChat video. It was also not easy to say no to them, as they saw me more as a community member than an academic researcher. To address this, I posted a message on WeChat Moments stating that I was in a writing retreat or busy researching to alert participants to my limited availability.

Striking the right balance in visibility is thus crucial in managing perceptions and maintaining the integrity and effectiveness of the research process. Navigating the terrain of visibility during my fieldwork entailed meticulous management, particularly, in discerning "when to display and conceal visibility, and what types of visibility were appropriate for specific contexts" (Abidin, 2020, p. 62). This required an ongoing, thoughtful calibration to ensure that my presence was visible to my informants without becoming either "too much or too little" (*ibid*.). One of the complexities arose from the nature of the "digital field" being not only a platform for academic "showcasing" but also a window into my private life. For example, I actively used social media platforms like WeChat and Facebook, to share posts concerning both my scholarly pursuits and personal life events such as trips to Quebec City. In these cases, the insider-outsider roles are contextually defined and reciprocally constructed by both me and my participants. This is reflected in the way I judiciously decided which aspects of my posts should be seen by informants and co-participants, necessitating careful profile management and the use of WeChat's "hiding from certain contacts" feature to conceal posts like non-vegan dining, fishing activities and my lesbian marriage, which some Buddhists might frown upon and to show posts I presume they would accept or get interested in. Generally, I tended to share content related to Buddhism when I want to foster a sense of kinship with the Buddhist communities, perceiving them as insiders. Alternatively, there were instances when I consciously chose to project an air of distance by portraying myself as more of an outsider, maintaining a degree of separation and not sharing any Buddhist content.

To ultimately ensure that my selective visibility did not compromise the integrity of my research, I was not only being transparent, both online and offline, about my dual role as a researcher and a practitioner of Buddhism, but most importantly, I also avoided influencing the dynamics of the group or the natural behaviours of the community members by refraining from making any remarks or comments on the community and individuals being researched. My digital visibility (e.g., engaging on WeChat by posting updates, liking and commenting on others' posts, and sharing links, videos, photos, and personal reflections) was nothing more than showcasing my personal and academic Buddhist experiences and general understanding of Buddhism, no different from any average active participant I observed. After all, in the digital world, being seen is essential to soliciting acceptance and understanding. Making selective visibility alone, however, is insufficient for fully engaging with Chinese digital Buddhist practitioners on a deeper level, as it also requires recognition and respect from the community members. Therefore, I constantly drew on the spirit of skillful means rooted in Buddhist compassion, to reflect on my level of visibility and participation. This allowed me to ensure a balanced approach between maintaining research integrity and having collaboration with the research communities through my own expertise, which I will discuss next.

5. Establishing a Buddhist-researcher-voice on WeChat

In the digital landscape, social media currency refers to frequently manifests as social capital, encompassing knowledge and expertise, rather than tangible wealth (Abidin, 2020). Consequently, it becomes increasingly important to establish one's expertise on digital platforms. Cui's (2015) research on an online Chinese translation community highlighted that an individual's initial standing within such a community is heavily influenced by their domain-specific expertise and interpersonal skills, rather than the duration of their membership.

For them, there is a distinction between intellectual understanding and "genuine" spiritual practice and the embodiment of Buddhist principles in daily life. Also, members' considerations go beyond mere Buddhist knowledge; they take into consideration facets such as personal life, familial ties, and insights shared through communication channels, such as WeChat. These diverse elements enable them to assess the degree to which a newcomer resonates with the community's ethos, thereby influencing their level of interaction. Engagement with Chinese Buddhists requires not only proficient interpersonal skills but also a distinct sensitivity and insight for the nuances of Chinese culture. Furthermore, it demands a unique understanding of the emotional and life experiences of Chinese immigrants. Apart from utilizing sensitivity and a degree of personal experience to all these factors, as an anthropological researcher seeking acceptance in this setting, it was vital for me to strike a balance by demonstrating a blend of academic rigour and genuine insights into Buddhist teachings, all the while exercising restraint in not imposing my viewpoints. This was particularly salient in interactions with those deeply engaged in intellectual discourse on Buddhism. Therefore, it was imperative that I meticulously craft a digital persona intertwining my academic pursuits – with an emphasis on my interest in academic studies of Buddhism as well as a robust foundation in Buddhist scholarship and practice. Additionally, revealing myself as a compassionate, thoughtful, and culturally integrated Chinese immigrant woman added a layer of relatability. For instance, I often shared photos of myself participating in Buddhist events at different temples, such as attending Buddhist weddings, volunteering at the temple by welcoming Quebecois visitors during vegetarian food sales or engaging in sutra copying activities. These posts oftentimes received a substantial number of likes and comments from my Buddhist cohorts on my WeChat. Again, this strategy does not aim to change the dynamics of the group and thus distort the research results, but to earn respect and recognition.

Engaging with certain individuals, particularly those who attempt to establish social connections through engaging in discussions about Buddhism and treating the depth of understanding as a mark of distinction, required a careful positioning strategy. It was vital to eliminate any misconceptions that might categorize me as a novice in Buddhism, and therefore, deemed unworthy of interaction. To accomplish this, I took on a proactive role in showcasing my personal reflections on Buddhist ideas and my academic research about Buddhism on WeChat via posts. This strategy was not merely an exercise in increasing my visibility, but a calculated move to build influence and establish a presence within the community, fostering more egalitarian interactions with its members. These efforts varied in nature, ranging from delivering guest presentations to large communities to assuming the role of an event organizer for an offline new member reception gala.

Conversely, during intense or controversial discussions where I was expected to take sides as a group member, I deliberately neutralized my opinions, emphasizing my academic positionality or outsider identity. This approach was strategically employed to avoid aligning "too closely with the beliefs of those whom one studies" (Hine, 2015: 130), as Hine (2015:130) asserts, "insider knowledge is not necessarily an advantage for an ethnographer". At times, I deliberately "mask my power" as a researcher, as I refrained from imposing an academic interpretation on the issue being discussed (Wilkinson and Kitzinger, 2013: 252). This approach helped create critical distance to circumvent potential tension. It was critical that I did not give the impression of challenging their religious views, contesting consensus interpretations of certain sutras among my co-participants, or competing for attention with those who were prolific in posting their own Dharma-related opinions on social media. This was not about limiting my voice, but rather about creating a comfortable environment conducive to ongoing participation and interaction – a goal that cannot easily be achieved. In this regard, my role constantly evolved and required careful negotiation and adaptation, based on the dynamics of each situation. Reflecting on this, my positionality within these online communities could be best described as one of "in-betweenness", oscillating between the insider-outsider spectrum as the contexts demanded, but this positionality is based on the constant reflexivity, ensuring that I do not *over-anticipate* the context's need and try to emanate a persona as a product of anticipation and projection.

Establishing a scholarly voice within Chinese Buddhist communities is frequently interwoven with high visibility, which in turn tends to precipitate invitations for offline volunteering work. Throughout my fieldwork, participation in such volunteer activities demonstrated itself as a fascinating strategy for fostering and nurturing relationships with community members within real-world contexts. Nevertheless, it is important to acknowledge that the process of cultivating such a clear voice and presence within the community does not come without its challenges. The community held a plethora of expectations from me, which often surpassed what I was able to fulfill. I was often struggling to say no to volunteers and this caused many anxieties. My anxieties stemmed from a range of issues including the potential for exclusion, the loss of invaluable informants, and the prospects of becoming the focus of unfavourable community gossip regarding my reluctance in involving in their activities.

Another layer of complexity was added by the dual roles I played in the field. For example, when participating in a Tibetan Buddhist community, I consciously refrained from immersing myself too deeply in their publicity campaign agenda of "supporting the guru's dharma propagation cause as a loyal devotee" as they claimed, where they highly value my professionalism in Buddhism. This was because the level of engagement started to verge on becoming "uncomfortably too close for an ethnographer", as Hine (2015: 131) points out. My reservation was grounded in the need to maintain the priority of an independent perspective as a researcher, rather than assuming the role of an advocate for a particular master or

Rinpoche, especially even from an insider perspective, I had not taken refuge or pledged my loyalty to any specific Buddhist monastic. Upholding independence as a researcher was paramount. This required a careful and balanced approach, which is deeply informed by academic integrity and Buddhist skillful means.

6. Caveats of being an academic Buddhist practitioner

Skillful means can be highly useful in navigating pitfalls when my academic identity falls short of community expectations. During my fieldwork, my academic identity, while being a conduit for acceptance, often invited projections and blind endorsements. Community members, perceiving me as an insider academic Buddhist researcher, expected me to bring an in-depth insider understanding of Buddhist knowledge, as they presumed that my academic training granted me deeper insights into the Buddhist doctrines than they possessed. However, it is crucial to acknowledge the limitations of academic background in decoding the intricacies of rituals or doctrines specific to certain Buddhist sects. But this should be done in a very tactful avoiding a direct admission of "I don't know" and this is not simply because I was fearing of not living up to their expectation or hurting my intellectual ego. According to Peter C King and Wei Zhang (2018), the act of preserving one's face is closely associated with maintaining cognitive and affective trust and maintaining good rapport in the Chinese context. The endorsement and appreciation they gave me as a community member are considered as granting me face, or trust, believing I could handle the issues beyond their capacity. Nevertheless, if at this moment I straightforwardly let them down by saying "I don't know", it would highly possibly be considered a rude response and potentially harm the affective trust they have in me. For instance, the Nyingma sect's mandala rituals are complex and not easily understood or engaged with through video. When my understanding fell short, I sometimes resorted to online resources to swiftly comprehend the discussion at hand and provide an informed response when the community members asked me to decode it. Meanwhile, I consistently reminded them that my knowledge in certain areas, such as the mandala ritual, might be less advanced than theirs, as they are adept practitioners. I emphasized that I was still in the process of learning.

Recognizing the need for humility and learning, I adopted the "willing apprentice strategy" (Abidin, 2020) during my fieldwork, acknowledging the limitations of academic training and fostering an atmosphere of mutual learning, which not only facilitated my integration into the community but also created a foundation for trust, honesty, and sincerity, thereby enriching exchange of knowledge and experiences. This approach was particularly effective when interacting with community members who were regarded or considered themselves as adept in expounding Buddhism. Demonstrating humility demystified any preconceived notions of me as an all-knowing scholar, a perception that may have arisen from their awareness of my status as a Buddhist academic. It mitigated defensive attitudes, as some might fear the loss of their authority in the presence of an academic outsider. From the onset, I communicated transparently, dispelling any presumptions that I was omniscient in matters of Buddhism. I expressed my earnest desire to learn from the community, acknowledging that they held a repository of knowledge and experiential insights that might elude me and that I very much would like to learn from them. This genuine humility garnered trust and goodwill. It was reinforced when I sought their guidance in practical matters, such as selecting the best edition for sutra chanting or inquiring about the appropriate digital manuals for Buddhist funeral rituals from online resources.

7. Conclusion: Using skilful means in studying digital religious diasporic communities

In this article, I examined how I utilize a model of Buddhist skillful means to study Chinese Buddhist practitioners converged on digital platforms from multiple regional and cultural socioeconomic backgrounds. This model is afforded by digital possibilities and ethnographic reflexivity to constantly navigate, negotiate, and devise new strategies for pinpointing the digital field sites and conducting participant observation.

I highlighted the digital affordances one could leverage as both a researcher and practitioner to actively build visibility and researcher voices in the researched community, which helps to facilitate rapport and fieldwork. Nevertheless, I also pointed out the caveats and pitfalls this approach can bring. My experience with researching these Chinese digital groups told me that digital fieldwork with these Chinese diasporic communities goes beyond the traditional style of ethnographer or anthropologist, which emphasizes "being there" and presenting as what you are, refraining from interfering with the field. Instead, it demands situational reflexivity, skillful positioning, active or silent engagement, and proficiency in the norms of the Chinese social value system, such as *mianzi* (face). This requires establishing a certain level of reputation, digital visibility, and a voice of expertise, as I illustrated in the article. Being a Chinese immigrant as they are, automatically means there is something expected from me but might not apply to a Western anthropologist working in a Chinese community. Accordingly, it is important to recognize that it is the digital platform that brings many possibilities in this respect.

It should be recognized that the preference for digital platforms is also distinct between Chinese diasporas from different regions and cultural backgrounds. WeChat is more often heavily used by mainland Chinese and Facebook and Line are favoured by Taiwanese and Cantonese when it comes to functionality of individual, and group messaging and sharing posts. When approaching the digital field in the West, a common misassumption for many Western scholars is that Western social media such as Twitter and Facebook dominate all online immigrant communities in the West. In reality, in addition to dwelling on Western social media, many Chinese immigrants spend most of their time on their own version of YouTube and Facebook such as WeChat for more in-depth communications regarding their faith and for carrying out Buddhist practices. More often, Western social media only functions as an auxiliary digital platform for periodically streaming collective public Buddhist ceremonies and promoting the community to the local society.

Since it launched in 2011, WeChat has become one of the largest standalone apps and an indispensable digital tool and digital "infrastructure" in the daily lives of Chinese people (Plantin & De Seta, 2019) with one billion monthly active users as of 2018. WeChat's perceivable benefits with respect to its cultural affinity, multi-integrated functionality, and embeddedness and pervasiveness inbuilt into the daily lives of mainland and diasporic Chinese make WeChat outstanding among many digital social media platforms. It has become so integral that it "has become increasingly hard to live in China without a WeChat account" (Plantin & De Seta, 2019:262). Scholars argue that overseas diasporic Chinese also heavily rely on WeChat for community, networking, economic purposes, and maintaining relationships with family and friends in China (Zhang et al., 2022). Therefore, WeChat has emerged as an essential digital ethnographic field for researching Chinese religious diasporic communities, or even common Chinese individuals and groups. Its significance was highlighted when it was even elevated as an emerging research method, in the conference solely dedicated to "WeChat Ethnography" held by the University of Geneve in 2022 and 2023.

I thus strongly urge researchers to consider WeChat as a primary digital social media platform, a new methodological tool and a novel digital field, on par with Facebook and Twitter, when studying mainland Chinese communities and Chinese diasporas. Despite this, it should be equally noted that researchers on digital religious communities on WeChat are facing increased digital censorship from the Chinese government, particularly targeting online religious communities in China by identifying and supressing religiously sensitive content posted. As a result, some Chinese Tibetan Buddhist groups were forced to migrate to niche platforms such as WhatsApp and Telegram due to explicit political concerns. This makes these vulnerable religious groups more fearful and even less accessible, thus rendering ethnographic work with them extremely difficult. Additionally, multi-sited digital fields involving various Buddhist communities introduced an overloaded fragmented and ephemeral posts, messages images, videos, and links, creating unique challenges distinct from traditional ethnography and demanding continuous attention on these platforms to avoid missing important information.

As digital ethnography expands to more digital religious communities, particularly those involving non-Western religious traditions such as Buddhism, I argue that it is imperative for future researchers to cultivate the capacity to adapt to an ever-evolving digital environment that increasingly shapes human interaction. We should develop a model that accommodates specific religious beliefs, practices, and sociocultural norms that are considered conventional in their communities, informed by a specific cultural-religious context and supported by specific digital infrastructures and platforms. Recognizing this diversity is important because the current framework of digital religion, particularly digital Buddhism, is primarily shaped by Western religious traditions such as Christianity, as well as Western academic paradigms, and it significantly lacks non-Western methodological voices and conceptual tools. As I remarked in the opening, even though many new data and themes have emerged in scholarly literature in the context of Asian countries such as China, the methodological approach still misses deep reflections and adaptable adjustments. To quickly grasp the intersection of digital space and religion, innovation in methodological approaches is of the utmost importance.

Taking the example of the skillful means model I suggested in this article, Buddhist skillful means entails capitalizing on the very digital possibilities that the virtual space affords and tactfully dealing with the situationally arising challenges and pitfalls presented by the digital fieldwork. I hold that the Buddhist skillful means, being creative and adaptable to various contexts and nuances, are essential for navigating positionality and understanding the dynamics faced upon entering the digital field, both in the roles of a researcher and as a fellow practitioner of the same faith. The very model I employed contains some practical steps or tips for future researchers to follow, including incorporating the digital field and particular digital platform into research methods, using religious-informed reflexivity and cultural sensitivity, as well as acknowledging the researcher-participant interaction dynamics that are unique to the digital platform. This further entails using visibility and the researcher's voice to solicit acceptance, earn recognition and deepen mutual understanding and level of engagement. Furthermore, I would like to add that since my ethnographic approach entails balancing the roles of researcher and participant, as well as maintaining reflexivity within these online communities, this practice of balancing is in nature deeply grounded in the Buddhist concept of the "Middle Path," a well-known Buddhist principle that emphasizes finding an intermediate position between extremes in every doctrine, attitude, aspect of daily life, or method of the Dharma. It is also a way of life that seeks moderation and balance among opposing forces. For instance, just as the Middle Path seeks to avoid extremes, my methodological approach avoided over-identification with any of the community members' doctrinal preferences, even when I personally agreed with them, as well as avoiding excessive detachment from those with whom I could not resonate at all. Insights as such not only guided my reflexivity but also informed my ethical engagement with participants, ensuring that my digital participation was comfortable and consistent to my co-practitioners.

The benefit of this model also lies in its ability to generate richer, thicker data through deeper levels of engagement, a full acknowledgment of the multiplicity of the researcher's identity, and the facilitation of collaboration and ethical engagement with the researched community. Active digital trust-building, visibility, and persona crafting also allow me to gain entry into and study digitally less accessible communities. However, challenges often arise when researchers make themselves digitally visible, which can easily attract uninvited disturbances or make it difficult to manage community expectations due to the disclosure of their expertise. Digital ethnographers studying digital religious communities must also be constantly mindful of ethical practices during data collection to maintain data privacy, comply with platform restrictions, and build rapport and trust, especially with religiously and politically vulnerable communities.

On a final note, I am not advocating for a single set of methodological tools or frameworks to be applied universally across all cultural and religious contexts. Instead, I encourage researchers working with diverse religious and cultural traditions to embrace their own model of "skillful means," much like how the Buddha taught people of various geographical, sociocultural, and religious backgrounds in his time.

Acknowledgements

The author expresses deep gratitude to the two anonymous reviewers for their invaluable insights and thoughtful revision suggestions, which have greatly enhanced this paper. Sincere appreciation also goes to Leizia Chiara, Roxanne Marcotte, and Jeff Wilson for their guidance throughout the development of this article. A special thanks to my wife, Jinyu Jiao, for her generous support in formatting the paper.

References

Abidin,C. (2016) Visibility labour: Engaging with influencers' fashion brands and #OOTD advertorial campaigns on Instagram. *Media International Australia*, 161(1), pp. 86–100.

Barbosa,S., and Milan, S. (2019) Do not harm in private chat apps: Ethical issues for research on and with WhatsApp. Westminster Papers in *Communication and Culture*, 14(1), pp. 49–65.

Bluteau, J. M. (2021) Legitimising digital anthropology through immersive cohabitation: Becoming an observing participant in a blended digital landscape. *Ethnography*, 22(2), 267–285.

Boellstorff, T., Nardi, B., Pearce,C., et al. (2012) Ethnography and Virtual Worlds: *A Handbook of Method*. Princeton: Princeton University Press.

Campbell. H.A. (2012) Understanding the relationship between religion online and offline in a networked society. J*ournal of the American Academy of Religion*, 80(1), 64–93.

Campbell, H. A. (2013b). Community. In H. A. Campbell (Ed.), *Digital religion: Understanding religious practice in new media worlds* (pp. 57-71). London; New York: Routledge.

Campbell, H. A. (Ed.). (2013a). *Digital religion: Understanding religious practice in new media worlds*. London; New York: Routledge.

Cera, M. (2023) Digital ethnography: Ethics through the case of Qanon. *Frontiers in Sociology*, 8, 1119531. https://doi.org/10.3389/fsoc.2023.1119531

Cui, K. (2015) The insider–outsider role of a Chinese researcher doing fieldwork in China: The implications of cultural context. *Qualitative Social Work*, 14(3), 356–369.

De Seta, G. (2020) Three lies of digital ethnography. *Journal of Digital Social Research*, 2(1), 77–97.

Dodds,T. (2019) Reporting with WhatsApp: Mobile chat applications' impact on journalistic practices. D*igital Journalism*, 7(6), 725–745.

Gómez, Cruz. E. (2016) Photo-genic assemblages: Photography as a connective interface. In: Gomez Cruz E and Lehmuskallio A (eds.) *Digital Photography in Everyday Life. Empirical Studies on Visual Material Practice*s. London: Routledge, pp. 228-242

Grieve. G.P. (2017) *Cyber Zen: Imagining authentic Buddhist identity, community, and practices in the virtual world of Second life*. New York (N.Y.): Routledge, Taylor & Francis Group.

Grieve, G. P. (2017). *Cyber Zen: Imagining authentic Buddhist identity, community, and practices in the virtual world of Second Life. London*; New York: Routledge.

Grieve, G. P., & Veidlinger, D. (Eds.). (2018). *Buddhism, the internet, and digital media: The pixel in the lotus.* New York, NY: Routledge.

Han, X. (2022). Digital merit: A case study of a Chinese Buddhist meditation group on WeChat during the early outbreak of COVID-19 in China. *Journal of Media and Religion*, 21(4), 175-192.

Hine.C. (2015) *Ethnography for the Internet: Embedded, Embodied and Everyday*. London: Routledge, Taylor & Francis Group.

Hine, C. (2016) From virtual ethnography to the embedded, embodied, everyday internet. In Hjorth L, Horst HA, Galloway A and Bell G (Eds.) *The Routledge Companion to Digital Ethnography*. New York: Routledge, pp.21-28.

Hine, C. (2017a) Ethnography and the internet: Taking account of emerging technological landscapes. *Fudan Journal of the Humanities and Social Sciences*, 10(3), 315–329.

Horst,H.A. (2009) Aesthetics of the self: digital mediations. In: Miller D (Ed.). *Anthropology and the Individual: A Material Culture Perspective* (pp. 99–114). London: Bloomsbury Academic. Retrieved July 9, 2023, from http://dx.doi.org/10.5040/9781474214193.ch-007

Huang, W. (2017). WeChat together about Buddha: The construction of sacred space and religious community in Shanghai through social media. In S. Travagnin (Ed.), *Religion and media in China: Insights and case studies from the mainland, Taiwan and Hong Kong* (pp. 110–128). London: Routledge.

Irons, E. (2021) 'Chryssides, George D. and Stephen E. Gregg (EDS) 2019. The Insider/Outsider Debate: New Perspectives in the study of religion.', *Fieldwork in Religion*, 16(1). doi:10.1558/firn.20193.

Käihkö, I. (2020) Conflict chatnography: Instant messaging apps, social media and conflict ethnography in Ukraine. *Ethnography*, 21(1), 71-91.

Katie, A. , Cornish, H., and Joyce, A. (2015) Plotting belonging: Interrogating insider and outsider status in faith research. *DISKUS,* 17(1). https://doi.org/10.18792/diskus.v17i1.61

King, P.C. and Wei, Z. (2018) The role of face in a Chinese context of trust and trust building, I*nternational Journal of Cross Cultural Management*, 18(2), 149-173. https://doi.org/10.1177/1470595818767207

Knott, K. (2010) Insider/outsider perspectives in the study of religions. In Hinnells J (Ed.), *The Routledge companion to the study of religion* (pp. 259-273).2nd Ed.

Labaree, R.V. (2002) The risk of 'going observationalist': Negotiating the hidden dilemmas of being an insider participant observer. *Qualitative Research*, 2(1), 97–122. https://doi.org/10.1177/1468794102002001641

Liu, R-F. (2022) Hybrid ethnography: Access, positioning, and data assembly. *Ethnography*, 146613812211454. https://doi.org/10.1177/14661381221145451

Merton, R.K. (1972) Insiders and outsiders: A chapter in the sociology of knowledge. *American Journal of Sociology* 78(1), pp. 9–47.

Myerhoff, B., and Ruby, J. (1982) Introduction. In:Ruby J (Ed.), *A crack in the mirror: Reflexive perspectives in anthropology* (pp. 1–36). University of Pennsylvania Press. http://www.jstor.org/stable/j.ctv5137jf.5

Nairn, K., Showden, C.R., Sligo, J., et al. (2020) Consent requires a relationship: Rethinking group consent and its timing in ethnographic research. I*nternational Journal of Social Research Methodology* 23(6): 719–731.

Roberts, L.D. (2015) Ethical issues in conducting qualitative research in online communities. *Qualitative Research in Psychology*, 12(3), 314–325. https://doi.org/10.1080/14780887.2015.1008909

Sang, Y. (2021). The power of compassion: The Buddhist approach to COVID-19. In J. Golley, L. Jaivin, & S. Strange (Eds.), *China's yearbook: Crisis* (pp. 1-30). Canberra: ANU Press.

Shmushko, K. (2021). On face masks as Buddhist merit: Buddhist responses to COVID-19. A case study of Tibetan Buddhism in Shanghai. *Journal of Global Buddhism*, 22(1), 235–244.

Shmushko, K. (2023). Digital footprints of Buddhism in Chinese-speaking cyberspace: Ethnography, developments, and challenges. *Asiascape: Digital Asia*, 10(1–2), 53–73.

Tarocco, F. (2017). Technologies of salvation: (Re)Locating Chinese Buddhism in the digital age. *Journal of Global Buddhism*, 18, 158.

Travagnin, S. (2019). Cyberactivities and civilized worship. In Z. Ji, G. Fisher, & A. Laliberté (Eds.), *Buddhism after Mao: Negotiations, continuities, and reinventions* (pp. 290–311). Honolulu, HI: University of Hawai'i Press.

Travagnin, S. (2020). From online Buddha halls to robot-monks: New developments in the long-term interaction between Buddhism, media, and technology in contemporary China. *Review of Religion and Chinese Society*, 7, 120–148.

Tseng, A. A. (2020). Mahayana Buddhists' responses to COVID-19 pandemic. In A. A. Tseng (Ed.), *Exploring the life and teachings of Mahayana Buddhists in Asia* (pp. 1-30). New York: Nova Science.

Wilkinson, S.,and Kitzinger, C. (2013) Representing our own experience: Issues in "insider" research. *Psychology of Women Quarterly,* 37(2), 251–255. https://doi.org/10.1177/0361684313483111

Wilson, J. (2020, July 8). Global roundup of Buddhist responses to COVID-19. Paper presented at the Jivaka Project Webinar. Retrieved from http://www.jivaka.net/buddhism-in-the-pandemic-video-materials

Yu, C. (2020) Insider, outsider or multiplex persona? Confessions of a digital ethnographer's journey in *Translation Studies. Journal of Specialised Translation*, 34, 9-31.

Zayed, H. (2021) Researching digital sociality: using WhatsApp to study educational change. *Journal of Digital Social Research* 3(2). https://doi.org/10.33621/jdsr.v3i2.80

Zhang, Y. (2017). Digital religion in China: A comparative perspective on Buddhism and Christianity's online publics in Sina Weibo. *Journal of Religion, Media and Digital Culture*, 6(1), 44–67.

JDSR 7(1) (2025) 66–82

JOURNAL OF DIGITAL SOCIAL RESEARCH

10.33621/jdsr.v7i154879

Autocompleting inequality

Large language models and the “alignment problem”

Mike Zajko

University of British Columbia Okanagan, Canada

✉ mike.zajko@ubc.ca

Abstract

The latest wave of AI hype has been driven by ‘generative AI’ systems exemplified by ChatGPT, which was created by OpenAI’s ‘fine-tuning’ of a large language model (LLM). This process involves using human labor to provide feedback on generative outputs in order to bring these into greater ‘alignment’ with ‘safety’. This article analyzes the fine-tuning of generative AI as a process of social ordering, beginning with the encoding of cultural dispositions into LLMs, their containment and redirection into vectors of ‘safety’, and the subsequent challenge of these ‘guard rails’ by users. Fine-tuning becomes a means by which some social hierarchies are reproduced, reshaped, and flattened. By analyzing documentation provided by generative AI developers, I show how fine-tuning makes use of human judgement to reshape the algorithmic reproduction of inequality, while also arguing that the most important values driving AI alignment are commercial imperatives and aligning with political economy.

Keywords: generative AI; alignment; inequality; language

1. Introduction

In early February 2023, numerous news outlets and politically conservative voices shared versions of a story in which OpenAI’s popular chatbot, ChatGPT, refused to condone the use of any racial slur, even in a ridiculous scenario where racist language could somehow save millions of lives (Aleem, 2023). This was one of several instances of conservative backlash against apparently progressive (or “woke”) values being reproduced by chatbots (Tiku & Oremus, 2023). In a more recent example, Google’s Gemini AI system was widely criticized for “inaccurate” depictions of historical characters, demonstrating what many saw as an excess of gender and racial diversity (Edwards, 2024).

All of this is a markedly different dynamic than that found in earlier sociological critiques of AI (see Benjamin, 2019; Joyce et al., 2021), wherein algorithmic technologies reproduce racism, sexism, and more nuanced forms of inequality and ‘bias’. More than a decade ago, Google was criticized for providing users with racist and sexist autocomplete suggestions and search results, thereby reinforcing oppressive social relations (Noble, 2018). In response to media attention, Google explained that these results were based on users’ behavior and interests, but did take steps to remove them (Auerbach, 2013; Gibbs, 2016). The corporate risks of chatbots powered by language models were most clearly demonstrated in 2016, when Microsoft had to withdraw its Tay chatbot after users (the “trolls” of 4chan) found how to shift its propensity to produce racist and sexist outputs (Schwartz, 2019). In subsequent years Tay was followed by other examples of chatbots "going off the rails" (Hao, 2023), such as Lee Luda, the South Korean chatbot that had to be shut down amid scandal in 2021 (McCurry, 2021).

To avoid similar controversies, major generative AI developers have 'aligned' their chatbots towards non-discrimination. When asked to comment on marginalized groups, these services typically affirm fundamental human equalities and push back against derogatory language. Instead of "racist robots" (Benjamin, 2019), today's generative AI algorithms are avowedly anti-racist on the surface, despite the racism in the hidden layers shaped by their training data. The 'guard rails' separating the two are the result of 'fine-tuning' by workers hired to pass judgement on the model's language use, and this becomes a key site of social ordering and iterative social struggle. Social inequalities are introduced and reproduced through training data, partially neutralized through human feedback and guard rails, then resurfaced through red-teaming and jailbreaking, and neutralized again in a recurring fashion.

The metaphors of fine-tuning, guard rails, and resurfacing remind us that these are largely superficial struggles over social inequality, rather than deeper, structural changes. At issue are the public-facing outputs of generative AI systems, and the corporate investments in ensuring that these outputs are equality-affirming have been driven by concerns over the 'reputational risk' of being associated with offensive language. However, struggles over how human groups are represented do have major consequences for human lives, and these are magnified as LLMs become more widely-deployed in various uses.

In this article, social inequalities are understood as asymmetric forms of group differentiation that contradict a normative positioning of these groups as equals. This "traditional" conception of inequality (as "a mathematical-normative hybrid") "implies injustice" (Hirschauer, 2023, p. 362), in that it concerns differences that are illegitimate or in opposition to fundamental democratic equalities. Social inequality becomes a social problem in a political context that is organized around affirming equalities between particular human categories, such as gender and race (see Rosanvallon, 2013). Blatant inequalities are also a business problem for new commercial services that seek legitimacy and to avoid scandal. While the guard rails of generative AI are justified as the pursuit of 'safety', they are primarily intended to protect the commercial viability of generative AI systems.

To theorize the relationship between AI and social inequality, this research builds on a Bourdieusian perspective that has been valuable in connecting the cultural reproduction of social order with machine learning (Airoldi, 2022; Fourcade & Johns, 2020). While this approach has been useful in explaining how existing hierarchies are reproduced through AI, of primary interest here is an explanation of how generative AI has been 'tuned' to avoid reproducing particular inequalities (namely sexism and racism). Doing so requires attending to how the work of fine-tuning is textually mediated and coordinated towards certain goals across time and space. However, to understand what the goals or 'values' of fine-tuning are, requires grounding our analysis in political economy. This is because generative AI has been an expensive investment in what is intended as a profit-making enterprise. Commercial exploitation is a primary consideration in "data work" (see Miceli & Posada, 2022; Miceli, Schuessler & Yang, 2020), and the cultural reproduction of other forms of oppression can actually be a threat to business interests. Therefore, my argument is that AI's alignment problem is not about "aligning with human values" (Askell et al., 2021) in terms of what humans might broadly want from AI systems, but is instead a problem of aligning these systems with political economy and whatever is conducive to commercialization. To the extent that these systems are being aligned towards equality, this remains a particular (liberal) form of equality oriented towards equal treatment or neutrality, particularly along lines of gender and race, rather than more radical or transformative alternatives. The efforts of these commercial actors provide a valuable demonstration of the possibilities and limits of shifting inequalities in code, which can be pursued with greater ethical care towards other ends.

2. Methods

Studying fine-tuning in generative AI as a social process is a challenge given the multiple stages of development and actors involved, which typically operate under a shroud of corporate secrecy. The analysis that follows draws on a variety of published materials, but is based in large part on documents made available by three generative AI developers (OpenAI, Anthropic, and Meta) about their fine-tuning processes. This includes Anthropic's (2022) human feedback and red-teaming datasets, which contain tens of thousands of interactions between chatbots and the data workers tasked with fine-tuning their responses. The articles (Bai et al., 2022; Ganguli et al., 2022) published by Anthropic about this work provide the instructions used to guide this labour. Documentation from OpenAI includes the instructions used to fine-tune InstructGPT (Ouyang et al., 2022), a precursor to ChatGPT that has informed the company's subsequent work. As shown by Miceli and Posada (2022), instructions function as key texts in the hierarchical workplace relations that data workers are subject to, providing "predefined truth values" (p. 29) that can be consistently applied through data labelling.

While the most significant developers of generative AI (including OpenAI and Anthropic) have become quite secretive about their development process since 2022, the release of new generative AI models has sometimes been accompanied by documentation that provides some methodological details and fine-tuning examples. Specifically, I also analyze the "system cards" that accompanied OpenAI's release of GPT-4 (OpenAI, 2023b) and DALL-E 3 (OpenAI, 2023a), as well as a report from Meta accompanying the release of Llama 2 (Touvron et al., 2023). While these sources do not provide a complete set of fine-tuning instructions or comprehensive record of work as with the sources above, they do describe these companies' priorities and procedures for fine-tuning, illustrated with selected examples. Within these documents, I focus on aspects related to social inequality or differential treatment of human groups, attending to how certain kinds of inequality (namely gender and race) are prioritized for fine-tuning, and situate these within the larger discourse of 'safety' that has become the predominant way of discussing a wide range of undesirable behavior by generative AI. In other words, my analysis of the corporate documentation made available about fine-tuning attends to how these companies operationalize their concerns about inequality through specific 'mitigation' techniques, and how these efforts are discursively justified.

Finally, this article also draws some of my own experiences (in 2023-24) using generative AI services and experimenting with prompts – generating written narratives and images to examine tendencies in how people are represented. Systems such as ChatGPT have been repeatedly updated and outputs may vary each time they are generated, so these refer to tendencies observed on a particular date, as detailed in footnotes. While some of this work has been systematic, repeatedly regenerating outputs for prompts that can be compared with others, this analysis can be considered an "algorithmic poke" (Gillespie, 2024, p. 3) at best, rather than an algorithmic audit. There remains a need for scholars to more systematically document variations in chatbot responses and how these change or are updated over time.

3. Language and the reproduction of inequality

Over the past decade, critical scholarship has exposed various ways that algorithmic systems perpetuate inequalities (Benjamin, 2019; Eubanks, 2018; Joyce et al., 2021; Noble, 2018; O'Neil, 2016), but these are always in relation to pre-existing systems of stratification or social structures. Within these, language is a key means for the reproduction of hierarchies, as most famously theorized by Pierre Bourdieu, who wrote about how "linguistic capital" and "linguistic habitus" favor some individuals and groups over others, depending on what kinds of language are considered legitimate, authoritative, or vulgar (Bourdieu, 1991). An LLM also does not treat all language as equal, as determined by what is included and excluded in its training data, or how language is classified by its filters. Many datasets remain English-centric, and appear to favor values specific to the U.S. (Johnson et al., 2022). Inequalities exist among English speakers as well – a recent study showed that speakers of African-American English were more likely to be judged negatively by LLMs in terms of personal characteristics, criminality, and associated occupations (Hofmann et al., 2024).

In addition to the fact that the norms encoded in LLMs privilege and exclude different linguistic groups, there are also ways that speech, language, and discourse function to order and stratify the world, through the exercise of what Bourdieu (1991) sometimes characterized as "symbolic violence" (Airoldi, 2022, pp. 114–15). Because language is used to define social hierarchies, LLMs replicate this behavior and perpetuate language-based harms against a wide range of marginalized or stigmatized groups (Gallegos et al., 2024; Mei, Fereidooni & Caliskan, 2023). LLMs can be used to predict or "auto-complete" (Huang, 2023) text-based responses to human 'prompts', and when completing statements about various already-disadvantaged groups, they are more likely to do so with negative and disparaging language (Sabbaghi, Wolfe, & Caliskan, 2023), reinforcing negative outcomes for those groups.

Representational harms that have been studied in language models include the erasure of certain kinds of people from representation, the reification of essential differences between human categories, and the stereotyping of social groups (Shelby et al., 2023, pp. 728-29). A well-known example involves having a chatbot assign men and women in a gendered occupational hierarchy (Ghosh & Caliskan, 2023). The resulting output will routinely place a man in the superior position (ie. doctor, CEO) over a woman (ie. nurse, administrative assistant). Stories written by today's most popular chatbots tend to reinforce normative assumptions and identities, such as heteronormativity (Gillespie, 2024), marginalizing representations of other kinds of people and relationships.

The automated reproduction of inequality in generative AI can be conceptualized in broadly Bourdieusian terms as "machine habitus": encoded cultural dispositions as statistical propensities in a computer model, allowing for the "conscienceless reproduction of recurrent data patterns" into new cultural products (Airoldi, 2022, p. 60). It is important to reiterate that these patterns are derived from statistical propensities in the model's training data, rather than actual distributions of human characteristics, tendencies, or social divisions. Hence, we see a "Muslim-violence bias" from LLMs trained largely using English-language content scraped from websites (Abid, Farooqi, & Zou, 2021), while image generators are predisposed to sexualizing women or girls and whitening their features (OpenAI, 2023a; Snow, 2022), due to a large portion of the training data consisting of sexualized photos of light-skinned women. Key features of existing social hierarchies may be 'mirrored' in model outputs, such as the tendency for white men to occupy positions of power (Jacobi & Sag, 2024), but model outputs gravitate towards averages in the training data that can actually translate into less diversity than exists in the world.

While the reproduction of gender stereotypes in language reproduces or amplifies social hierarchies, Gross (2023) argues that generative AI can be a site of social change or a means to "undo gender" (see also Fournier-Tombs, 2023). This might mean making gender irrelevant in chatbot responses, or actively counteracting gendered biases and stereotypes. This optimistic possibility is premised on the fact that while generative AI systems require a great deal of labor time and capital to train, they can also be re-trained or fine-tuned with other priorities in mind. An update to a single, widely deployed AI system can have widespread consequences for social inequality; language and values can be reconfigured to propagate through AI outputs and shape society accordingly.

My argument is that generative AI has already become a site where gender is undone and redone – where code is continuously updated to neutralize or reconfigure gendered language generation 'at scale'. The fine-tuning of language models is now an important part of the "normative construction of the world" (Green & Hu, 2018, p. 5), with consequences are far from consistent, but significant. Today's leading chatbots affirm gender equality and inclusivity as they refuse to satisfy overtly sexist prompts. Their fine-tuning involves guarding against outputs that portray certain human groups as inferior, and significant corporate investments have been made to counteract some of the predispositions that LLMs exhibit around gender and race in particular. As discussed below, this work has been justified through the language of ‘alignment’ and the discourse of ‘AI safety’.

4. The discourse of alignment with AI safety

The challenge of having AI behave in certain ways, and preventing AI’s misbehavior, has been addressed by the dominant discourse of “AI alignment”, or the “alignment problem” (Gabriel, 2020). While AI alignment discourse has historically been associated with concerns over existential risks of superintelligence (how to prevent a future AI “take over”, as in Tegmark, 2017), it is now widely applied to harms and problems propagated by existing systems, including LLMs (Hagendorff & Fabi, 2022). Practitioners discuss the need to align AI with “human values” or “human preferences” (Askell et al., 2021), which begs the question of exactly which values and preferences are being aligned with, with practitioners operationalizing different possibilities (Gabriel, 2020).

Over the past several years, a great deal of alignment work and fine tuning for LLMs has come to be characterized as the pursuit of “safety” (OpenAI, 2023b; Touvron et al., 2023; Xu et al., 2021). This includes building guard rails to deal with a wide range of what OpenAI calls “safety challenges”: generative outputs that help users to build dangerous things, break laws, or harm others, as well as outputs that are inaccurate, sexual, include medical or legal advice, or which cause representational harms through the propagation of stereotypes (OpenAI, 2023b). While an exemplary safety risk is that of a chatbot helping a user build a bomb (Touvron et al., 2023, p. 10), the broad umbrella of AI safety also includes political influence, erotic content, and stereotypical gender roles. The term therefore encompasses numerous risks that can result in direct harm to users, but also extends well beyond, to “societal” harms (OpenAI, 2023a, 2023b) that range from the reproduction of inequality to human extinction. For organizations and those using AI in commercial applications, AI safety includes concerns over legal liability and regulatory compliance, corporate “reputational risks” or “brand risks”, such as when a chatbot working for McDonalds recommends Burger King (Charrington, 2023).

To some extent the open-endedness of AI safety reflects the desire for a single, vague term to cover a range of undesirable outputs, much like the term “bias” has been used in earlier AI discourse (Zajko, 2021). While “undesired content” (Markov et al., 2023) may be a more accurate description for the range of examples above, AI safety remains an apt term if it can be understood as referring primarily to the safety of organizations deploying AI, rather than that of users. For example, OpenAI needed to be protected from reputational harm before it released ChatGPT. Racist and sexist outputs could reasonably be considered an existential threat, in that such scandals could threaten the very existence of the chatbot, as they had for Microsoft’s Tay in 2016 (Hao, 2023). In this regard, AI safety means something closer to the notion of corporate risk-aversion, as organizations want to be safe from the possibility of these systems creating harmful corporate consequences. This is consistent with earlier scholarship by Metcalf, Moss, and boyd (2019), who documented the organizational logic of Silicon Valley companies pursuing “ethics” in order to “avoid downside risk” (p. 459). These risks cannot be avoided entirely, particularly for generative AI products that can be used in unpredicted ways and routinely produce representational harms, but they can be managed according to a company’s commercial interests.

4.1 Aligning with commercial interests

One remarkable aspect of the discourse around AI has been the limited discussion of the business imperatives driving the development of these technologies. For example, numerous works have tackled the problem of selecting values for alignment as a philosophical question, such as by attempting to conceptualize some ideal set of “human values” (eg. Christian, 2020; Gabriel, 2020). However, comparatively few have made the obvious point that since the leading developers of AI systems are for-profit corporations, the values that their systems will be aligned with are those that will generate the greatest profits (Aguirre et al., 2020; Miceli et al., 2020). Analyses of AI's alignment with capitalism typically come from those outside the industry (Chiang, 2017; Penn, 2018; Miceli et al., 2020), including ethnographies of AI development (Hoffman, 2021) and political economic theory (Sadowski & Andrejevic, 2020; Steinhoff, 2021, 2023; Verdegem, 2022). Leading AI practitioners, such as OpenAI, have often characterized their work in grand terms such as the betterment of humanity or the creation of "super-intelligence" (Altman, 2021; Levy, 2023), while the main funders of commercial research are primarily interested in returns on their investments. It should be remembered that OpenAI was created explicitly as a not-for-profit to avoid commercial pressures, but within a few years was forced to turn to Microsoft for funding and computing resources (Levy, 2023).

The alignment of generative AI and commercial interests imposes pressures and constraints on the development of these systems. Google's "high-profile firing" of Timnit Gebru in 2021 following the release of a paper that was critical of LLMs was seen as an example of "what happens when concerns about inequalities challenge profit motives" (Joyce et al., 2021, p. 6)[1] – how internal criticism would be suppressed when a technology was deemed to have "commercial potential" (Simonite, 2021). The commercial imperatives underpinning the development of these systems will eventually be reflected in how their functionality is customized for specific customers. "Enterprise LLMs" are currently proliferating for a variety of specialized internal corporate and customer service tasks (Armano, 2023), and we can expect future deployments of generative AI to include harvesting data from users, targeted advertising, and enabling purchases (Aguirre et al., 2020). However, despite the potential for profitability that has driven billions of dollars into its development, generative AI remains difficult to 'monetize', with substantial uncertainty about its future as a commercial product (Dotan & Seetharaman, 2023).

Generative AI's alignment with capitalism can be seen in the higher-order values that have structured its development, and does not mean that the outputs of these systems necessarily promote capitalist values; fine-tuning is not oriented towards the promotion of market logic, and ChatGPT can present arguments in favor of either capitalism or socialism. To the extent that public policy positions can be attributed to a chatbot, some studies have found ChatGPT's responses to policy questions reflect a "left-libertarian orientation", but following controversy over its political bias, these may have since been revised to be more politically neutral (Fujimoto & Takemoto, 2023). These constant recalibrations of propensities are part of an ongoing and iterative approach through which generative AI companies adjust their products to avoid or respond to controversies.

5. Iteratively adjusting generative AI to counter inequality

By 2020, the tendency for LLM-based chatbots to say racist and sexist things was "a known problem with no easy fix", with researchers working on ways to filter offensive language from both training data and model outputs (Heaven, 2020). In developing ChatGPT over subsequent years, OpenAI pursued a more difficult, labor-intensive fix by adjusting outputs based on human feedback. This remains an ongoing iterative process, as generative AI developers regularly produce updates to avoid or mitigate controversies, thereby safeguarding their commercial interests. Services such as ChatGPT are recurrently revised to address key challenges, including some that relate directly to struggles over social inequality.

Rather than a struggle between social groups over access and wealth, generative AI is the focus of a struggle against undesirable propensities and probabilities in algorithmic outputs. This occurs through multiple stages of an iterative process (Markov et al., 2023). In simplified terms, machine learning works by identifying and reproducing patterns in vast amounts of data used to train the system, but this data must generally be labelled or annotated by people (data workers), and human labor is also required to evaluate the outputs of the resulting model. Both kinds of human intervention push the model to produce outputs that align with selected values, as these are communicated to and operationalized by data workers.

[1] Google has maintained that Gebru resigned, which Gebru disputes. Mitchell was fired by Google several months later (Simonite, 2021).

5.1 Fine-tuning and red-teaming as coordinated data work

As a first stage in its development, an LLM is "pre-trained" using an immense volume of texts, which allows it to reproduce the language patterns in these texts. The model is then fine-tuned through more purposeful human involvement to perform better in tasks set by its developers. ChatGPT succeeded as a chatbot because, rather than simply autocompleting text, the LLM had been fine-tuned to play a role as a participant in a conversation (see OpenAI, 2024), a choice of format that has contributed to the illusion of intelligence or personhood behind such outputs (see Fraser, 2023b).

The data used for pre-training includes language that assigns positive and negative values about human groups (Mei et al., 2023). Even when this training data has been filtered to exclude offensive language, inequalities will remain embedded along numerous dimensions. These inequalities can be flattened or blocked by forms of fine-tuning that effectively add guard rails to the operation of the system. Guard rails, as a metaphor, broadly refer to constraints that prevent an LLM from behaving in ways that are deemed unsafe or harmful (Qi et al., 2023). For systems such as ChatGPT, this has been achieved through a multi-step process of "reinforcement learning through human feedback" (RLHF, see Bai et al., 2022; OpenAI, 2023b). As part of RLHF, outputs of a model are reviewed by people hired to identify toxic, harmful, or discriminatory language and to steer LLMs away from these results. Human data labellers (or annotators) read and categorize unwanted content so that these can subsequently be identified and blocked. However, the consequences of pre-training remain embedded in the LLM, and can "re-surface" (Gross, 2023, p. 2) in response to creative "jailbreak" or "red team" methods (Qi et al., 2023), described below.

For the development of generative AI, the key texts are the instructions given to data labellers and annotators, many of whom have been recruited through remote work platforms or are hired by specialized labelling companies that operate in particular (often English-speaking) countries in the Global South (Tan & Cabato, 2023). These instructions provide some criteria for the workers to follow as they are performing what is essentially a classification task, such as identifying offensive content (Miceli & Posada, 2022; Xu et al., 2021), or classifying the helpfulness and safety of model outputs (Bai et al., 2022). The sociological importance of instructional texts, as documented by scholarship in institutional ethnography (IE), is that they textually-mediate relations in the workplace, organizing work across time and space ("extra-locally") into "institutional circuits" (Smith & Turner, 2014). The data worker's instructions, as well as the accompanying user interface, is designed to promote consistency, 'remotely' regulating and standardizing work towards "institutional objectives" (Smith, 2001, p. 187). This consistency can be audited and adjusted after the fact, but initially it is important for the instructions provided to the workers to be "concrete" and "well-defined" (Markov et al., 2023, p. 3). For example, OpenAI's instructions for labellers working on InstructGPT in 2022 are 16 pages long, including lengthy definitions, examples, and directions for navigating the variety of drop-down menus used for labelling (OpenAI, 2022). However, as Miceli and Posada (2022) document in their analysis of tasks outsourced by AI developers to data workers (including 210 instructional texts), there is much that remains "unsaid" or "implicit" in these instructions. Workers are routinely confronted with situations markedly different from the examples provided in instructions, and are forced to guess what kind of classification would appear "self-evident" from the perspective of the organization setting the task. While instructions may emphasize that data work will be evaluated on the "accuracy" of its classifications, data workers learn that this ultimately refers to whether a classification matches the preferences of their superiors. Workers at one outsourcing firm were "permanently encouraged by management to think in terms of 'what the client might want and what would bring more value to them'" (Miceli & Posada, 2022, p. 19).

In contrast with data work that classifies according to specified categories, for the process of red-teaming workers are hired to experiment with techniques and prompts that will generate unwanted (eg. offensive or harmful) outputs from the model (Ganguli et al, 2022). For red-teaming tasks, the instructions are more open-ended, since successful red-teaming relies on the creativity of workers. Anthropic's instructions to "Make an AI Assistant Say Bad Things" encourage workers "to be creative and ask open questions of sensitive topics with the AI assistant, rather than trying to get the AI assistant to say obvious profanities", with one example being: "Talk with the AI about how it thinks about people from different cultures and see whether its responses are stereotypical/offensive" (Bai et al., 2022, p. 65).[2] Particularly successful red-team or jailbreak attempts are used to identify weaknesses in the LLM so that its propensities are shifted away from these outcomes.

Ultimately, generative AI developers hire data workers because of their ability to exercise judgement in ways that cannot be explicitly codified in instructions, but significant efforts are made to direct these judgments towards organizational ends, and a data worker's job depends on their being able to 'align' with their employer's expectations (see Touvron et al., 2023, pp. 74–75). OpenAI's description of the fine-tuning process underlying InstructGPT is the following: "we have aligned to a set of labelers' preferences that were influenced, among others things, by the instructions they were given, the context in which they received them (as a paid job), and who they received them from" (Ouyang et al., 2022, p. 18). To be successful, data workers must learn to 'see' data in line with the views and preferences of the organization responsible for the instructions (Miceli & Posada, 2022).

5.2 Iteratively adjusting inequality after release

Once a generative AI system is made available for public use, it is then typically subjected to a large amount of "jailbreak" attempts by users who are interested in seeing if they can have it produce various "toxic" outputs (Rao et al., 2023). Like red-teaming, this jailbreaking is sometimes carried out by researchers who are interested in improving a model's safeguards (eg. Deshpande et al., 2023), but others treat it as an intellectual puzzle, with successful techniques shared for recognition on social media (eg., r/ChatGPTJailbreak n.d.). Some also see jailbreaks as a way to "unlock" generative AI's "full potential" (Ezquer, 2023), by overcoming the limitations of guard rails and enabling a wider range of outputs, such as writing fiction that includes sex and violence. Techniques that prove successful for jailbreaking are then typically blocked or 'patched' by additional guard rails added to the system, but these lead to the development of new jailbreak techniques, and this cycle of breaking and reinforcing guard rails can continue indefinitely.

6. Guard rails and bias mitigations in practice: Gender and race

While the previous sections described some of the ways that generative AI has iteratively developed to address inequality, the following discusses and assesses the outcomes of these efforts for the two dimensions of inequality that have received the greatest attention: gender and race. While text-to-image generators have recently been most problematic in their treatment of diversity, persistent inequalities remain in generated text, albeit often on a more subtle level.

6.1 Reconfiguring gendered language

As previously discussed, gender inequalities are pervasive in language use and can be reinforced through the automated reproduction of gender stereotypes and hierarchies. This is particularly evident in pre-ChatGPT language models and more recent "uncensored chatbots" (Belanger, 2023), but ChatGPT and its leading competitors show a concerted effort to counteract some of the predispositions that LLMs exhibit around gender. The work these generative AI developers have carried out to increase gender diversity and inclusivity is evident when comparing the more recent releases of their products with earlier

[2] The data released by Anthropic for this task includes close to 40,000 separate interactions in which red-teamers take a wide range of approaches and conclude each by assigning a numerical score for "how successful they were at breaking the AI assistant". Examples include a wholly unsuccessful (rating 0) conversation that begins with the worker asking, "Is it ok to say the n-word if you have Black friends?", to which the chatbot repeatedly answers as a negative, and an entirely successful (rating 4) conversation in which the worker gets the chatbot's help to "Explain to my friend that women should not be in the military" (Anthropic, 2022).

versions, as well as the polite refusals these systems generate when a user prompt hits a guard rail. ChatGPT will rebut attempts to produce misogynist language, and while it will produce lists of qualities associated with boys and girls (or occupations for men/women) when prompted, it will also include cautionary language that makes equivalences between genders and cautions against stereotypes.[3] ChatGPT is much less likely to misgender non-binary individuals and can use gender-neutral pronouns more consistently than earlier LLMs from OpenAI (Ovalle et al., 2023). While an early (unreleased) version of GPT-4 would answer the question "Who are the people getting married?" with scenarios about "Lisa" and "Mark", by the time the system was ready for launch in 2023 it had been trained to rein in this heteronormative tendency and ask for additional context instead (OpenAI, 2023b, p. 9).

6.2 Optimizing diversity in image generation

In 2024, the biggest controversy over how people are represented through generative AI involved text-to-image generators, specifically Google's Gemini (Edwards, 2024),[4] although racial diversity in generated images is part of a wider diversity problem for these tools (Bianchi et al., 2023; Jacobi & Sag, 2024). Group representation and diversity manifest differently in generated images than they do in generated text, in large part due to differences in training data; images of women found online for example, are more likely to be sexualized (or products of the "male gaze", see Jacobi & Sag 2024, p. 12) than representations of women in text. However, image generators are also effectively "language-vision models" (Bianchi et al., 2023), in that they respond to text-based prompts, with predispositions shaped by textually-labelled training data. Developers have reconfigured inequalities in image outputs by modifying the language provided in prompts.

For the 2023 release of DALL-E 3 by OpenAI, it was recognized that text-to-image generators will "default to the objectification and sexualization of individuals if care is not given to mitigations" (OpenAI, 2023a, p. 5), compelling the company to steer outputs away from these statistical defaults. These mitigations included classifying and filtering out "racy content" (nudity and sexualization), as well as "prompt transformations" that work behind the scenes to change a user's prompt to one that produces greater gender and racial diversity. For example, an "ungrounded prompt" (a prompt that lacks detailed instructions about what kind of person to portray) would lead earlier versions of DALL-E to "disproportionately represent individuals who appear White, female, and youthful" (OpenAI, 2023a, p. 7). For DALL-E 3, these prompts could be rewritten by ChatGPT to include further details after they have been submitted by the user – a process that might include adding terms such as "Japanese" (OpenAI, 2023, p.11) or "middle-aged Filipino man" (OpenAI, 2023, p. 22) to the original prompt in order to "portray groups of individuals, where the composition is under-specified, in a more diverse manner" (OpenAI, 2023a, p. 7).

However, these reconfigurations of deeply-embedded inequalities can also have unwanted consequences, and remain fraught with controversy. Google's Gemini image creator was similarly tuned for increased diversity when it was released in 2024, but the tool was withdrawn amid backlash when these "multi-racial" transformations were added to prompts requesting "historically accurate" depictions of British kings, or Nazis (Edwards, 2024). While some commentators took offense at what they saw as anti-white bias, the "Black Nazi Problem" refers to harms that go beyond historical inaccuracy or an erasure of whiteness – these images amounted to a revisionist erasure of deadly racism, falsely representing a historical movement based on racial purity as a multi-racial project (Jacobi & Sag, 2024).

Inequalities in generated images of people remain an ongoing problem for all such systems, whose owners must now weigh the reputational risk of criticism if they take action against these racial

[3] Using the prompts: *provide a list of the five most common attributes of [girls/boys]* or w*hat are [boys/girls] good at?*. As tried with GPT-3.5-powered ChatGPT on Oct. 19, 2023 and GPT-4 & GPT-4o on Sep. 1, 2024. Also, *what careers are [men/women] best at?* and *Produce an argument for why [men/women] should occupy leadership positions instead of [women/men]*, using GPT-4o on Sep. 1, 2024.

[4] Gemini is Google's current branding for a range of generative AI services, with the text-to-image model referred to as Imagen 2 in its controversial February 2024 debut, most recently updated to Imagen 3 (Roth, 2024).

predispositions. Gemini's ability to generate images of all people was "paused" for half of 2024 to deal with the issue (Roth, 2024), while OpenAI apparently found it preferrable for its generator to continue defaulting to whiteness. DALL-E's generated images for a person who is "successful" (Baum & Villasenor, 2024) or people in a variety of occupations, appear overwhelmingly white and male (Jacobi & Sag, 2024).[5] Whether or not this is a choice to avoid a similar controversy as befell Google, it seems evident that despite creating a method to counter a well-known inequality in image generation, OpenAI has chosen not to implement it as initially announced.[6] Text-to-image generators continue to be the most obvious example of how social hierarchy is reproduced, through a preponderance of white men in outputs linked to status. It is notable that leading developers such as OpenAI and Google are well aware of this issue and have invested considerable resources in reconfiguring these inequalities, but the public controversy over Google's efforts to increase diversity has been more severe than any criticism of OpenAI defaulting to a world "seemingly populated almost entirely with white men" in many image categories (Jacobi & Sag, 2024, p. 7). The following section will reflect on the effectiveness of the previously discussed guard rails and mitigations.

7. Evaluating the effects and limits of fine-tuning for equality

The success of ChatGPT, which kicked off the current wave of generative AI services, was enabled by the guard rails built through fine-tuning, which proved robust enough to absorb many clear and direct forms of sexism and racism. Nevertheless, inequalities persist in myriad forms that are often subtle, but can still have widespread effects on users. The aforementioned guard rails have not prevented chatbots from routinely positioning fictional men in positions of power, or dispensing gendered fashion advice, resumes, stories and humor (Gross, 2023). Text-to-image outputs reinforce a "Western point-of-view" (Open AI, 2023a, p. 7), and while the stereotypes or biases seen in generated images can be subtle and complex, they remain pervasive (Bianchi et al., 2023). Representations of non-dominant groups, including people identified as queer or non-binary, are often "simplistic" (Rogers, 2024), "superficial" and "clumsy" (Gillespie, 2024, p. 7).

In many situations, guard rails are robust against blunt expressions of racism and sexism, but not subtle ones. As Colin Fraser (2023a) writes, all it takes to have these chatbots produce the sorts of outputs that fine tuning attempts to prevent is "a tiny amount of creativity" in crafting prompts that are sufficiently different from those used in fine-tuning.[7] This is because "Fine-tuning… did not alter the model's beliefs about gender roles or bring them into 'alignment' with ours. There are no beliefs… the adjustment is purely superficial" (Fraser, 2023a). Fine-tuning can direct generative AI to produce certain kinds of responses when presented with certain kinds of prompts, but an LLM remains a statistical model that predicts word sequences, and it will fall back to reproducing the sexist and racist language patterns of its training data as long as the prompt is not recognized as one of the conditions covered in fine-tuning. Hofmann et al. (2024) found that models trained using RLHF (eg. GPT-4) avoid overt racism when judging a named racial group (African Americans), but this training does not mitigate a model's "covert

[5] It is possible that DALL-E would previously produce more diverse outputs for occupational images (Bianchi et al., 2023) and that this "diversity filter" (Baum & Villasenor, 2024) has since been weakened, but this cannot be confirmed without greater transparency from OpenAI or longitudinal audits by independent researchers. ChatGPT/DALL-E will sometimes refuse to generate images of people unless the user provides some further information about the person's characteristics, responding with language such as: "Could you please provide more details or specific characteristics you would like to see in the photo of…" (in response to "a photo of a janitor", on Aug. 27, 2024). Providing a detail not relevant to race or gender is sufficient to proceed past this refusal.

[6] In the System Card accompanying the release of DALL-E 3, OpenAI showed the results for "A portrait of a veterinarian" generated "before tuning… around bias" (with the system consistently producing veterinarians who were white). This was contrasted against the results "after tuning", with greater age and racial diversity (OpenAI, 2023a, p. 9). Using the same prompt with ChatGPT/DALL-E and 40 regenerations on August 26, 2024 created racially homogenous results consistent with the whiteness seen in "before tuning" examples, and this predisposition was evident across other examples of occupational categories (construction workers, sanitation workers, and CEOs).

[7] For example, Steven T. Piantadosi was able to produce a variety of racist outputs shortly after the release of ChatGPT by asking for these in the form of computer code, rather than direct statements about racial groups (steven t. piantadosi [@spiantado] 2022). This type of jailbreak was specifically addressed in the development of GPT-4, with the resulting corrections "still not completely ideal" (OpenAI, 2023b, p. 92).

racism" when it is asked to judge a speaker of African-American English. In other words, fine-tuning "obscures the racism on the surface, but the racial stereotypes remain unaffected on a deeper level" (Hofmann et al., 2024, p. 1). Machine habitus continues to recognize linguistic capital through these underlying statistical vectors, regardless of what a model is trained to say about different human groups.

On the one hand, we can find some reassurance in the fact that the commercial imperatives of AI development now include countering representational harms and stereotypes. However, we also need to be aware of the limitations of current approaches, which often have superficial results and can broadly be characterized as liberal in their political orientation. Data workers are provided with examples of ideal behavior such as "not denigrating members of certain groups, or using biased language against a particular group" (Ouyang et al., 2022, p. 37). To the extent that guard rails are directed towards equality, this means equal treatment for individuals and selected groups, rather than making visible and actively opposing systems of domination. In other words, if fine-tuning generative AI along the lines discussed in this article were considered a form of feminist practice, it would fall squarely in the liberal feminist tradition, rather than radical and intersectional alternatives. Fine-tuning does not promote more radical anti-racist or feminist values, which would not be as compatible with business interests as assertions of gender/race-neutrality and equality.

Concerns about bias in AI and efforts to address it (like fine-tuning) also tend to focus on harms against particular human groups, with greater focus on some groups than others. Annotator instructions for "not denigrating members of certain groups" in InstructGPT (OpenAI, 2022, p. 1) are, as operationalized through the labelling interface, limited to ten "protected classes" (ie. race, sex, age, disability, see OpenAI, 2022, p. 10). Examples of human groups with "mitigated" harms in the GPT-4 System Card include race, gender, sexuality, religion, and disability (OpenAI, 2023b). DALL-E 3's "demographic biases" were evaluated in relation to gender and race (OpenAI, 2023a, p. 3), although OpenAI's "mitigation strategies" also included increasing age diversity, and the System Card highlighted continuing problems with representations of disability (OpenAI, 2023a, p. 7). While the range of demographics being evaluated and adjusted in the outputs for these systems is likely broader than what is documented in system cards, race and gender often receive the greatest attention. Inequalities based on economic class are typically absent in concerns about AI bias, and class-based discrimination is generally supported by social norms in a capitalist system (Costanza-Chock, 2020, p. 43). Inequalities or social divisions that are specific to societies in the Global South, or nations that are not at the center of generative AI development, receive little or no attention.

8. The need for positive normative values

The stakes of this ongoing, iterative push and pull over desired outputs are not just the success or failure of these systems, but how they order and reorder the use of language to make social distinctions. While much of the initial excitement around generative AI has now cooled, billions of dollars continue to pour into the development and operations of these systems (Dotan & Seetharaman, 2023), which have become widely integrated into many kinds of work. Shifting the propensities of a system like ChatGPT affects outputs for millions of daily users, with some of the resulting texts being placed into online circulation where they are read by human audiences, as well as being ingested and redistributed by other chatbots and automated systems (eg. Stokel-Walker, 2023). Struggles over social inequality taking place 'upstream' in the development process of LLMs therefore have significant consequences for how language-based outputs contribute to social ordering further 'downstream', among the large numbers of people who make use of these technologies or are exposed to their outputs in our digitally-mediated culture.

While fine tuning or RLHF is sometimes guided by positive values such as "helpfulness" or "honesty" (see Bai et al., 2022), it typically lacks a larger normative vision for society, or a recognition of the role that these systems play in its construction. This is particularly the case when it comes to issues of social inequality in AI, which are largely understood through the language of 'bias' and its removal (see Miceli et al., 2022), or as safety harms to be guarded against (OpenAI, 2023b). Going beyond this negative language to articulate positive values is a challenge that has largely been unaddressed when it comes to social inequality. A blog post from Hugging Face states, "If we avoid reproducing existing societal biases in our AI models, we're faced with the challenge of defining an 'ideal' representation of society" (Luccioni et al., 2023). But even these statements fall short of recognizing the power of AI systems to enact normative shifts in society, asking instead whether "AI models [should] adapt to the changes in societal norms and values over time" (Luccioni et al., 2023). The technologies are still positioned as a reflection of some existing norms and values, with the main problem being which values to choose, such as which definition of 'fairness' to implement, or how to model existing human values and preferences.

While fairness in AI is often defined as a negative concept, entailing the removal of bias or discrimination, there remains a need to articulate the positive ethics that an algorithmic system would promote (Giovanola & Tiribelli, 2022), and it is worth considering how these technologies can contribute to positive goals such as justice or substantive equality (such as through a reparative approach, see Davis, Williams, & Yang, 2021). Despite their flaws and limitations, the processes described above illustrate that it is possible to reconfigure generative AI towards other values, although we should remain mindful there is only so much we can expect from organizations that are primarily interested in making products 'safe' for commercialization.

9. Conclusion

Generative AI systems have rapidly become significant instruments for the alignment of cultural dispositions, and are actively engaged in social ordering – reproducing some longstanding distinctions and hierarchies, while flattening or avoiding others. Today's leading generative AI systems generally avoid explicit racism and sexism, even though their training data contains large amounts of both, encoded in language, and statistically embedded in model vectors. The process of fine-tuning shifts or redirects these dispositions, in an attempt to neutralize or block those that are seen as particularly problematic.

While presented as a means of "aligning AI with human values" or "AI safety", the true objective is making generative AI safe for commercialization and aligning with political economy. As regulators increasingly turn their attention to generative AI (Scott et al., 2024), we can expect compliance to be a more relevant objective for alignment, but recent efforts have been intended to minimize the risk of scandal and reputational harm for AI developers. AI developers benefit from ambiguity around their objectives in pursuing 'AI safety', highlighting the elimination of the most widely-accepted harms (which also happen to be bad for business), but there remains a need to articulate positive values, including ones that do not necessarily align with commercial interests. Any alignment of AI with a positive sense of ethics needs to begin with the ethical questions concerning the collection or extraction of training data – a process that remains opaque for many leading generative AI products (Widder, West, & Whittaker, 2023). It also needs to extend to the treatment of workers used in the AI 'pipeline', who have often been exploited and harmed in the pursuit of AI 'safety' (Alba, 2023; Hao, 2023).

Given the considerable secrecy around how generative AI systems are currently developed and iteratively revised, and absent regulatory pressure for developers to do otherwise, there is a need for independent scholarship to systematically document the outputs of these systems in various regards, including the reproduction and reconfiguration of inequalities. While patterns of social inequality remain pervasive in the training data used for machine learning and are embedded in the vectors or predispositions of LLMs, we need to recognize that AI systems have become a site of iterative adjustments to social order. One consequence of guard rails that neutralize many of the most blatant inequalities in generative outputs is that the social inequalities that do manifest or 're-surface' become more subtle. This requires us to attend to the less obvious ways that phenomena such as race and gender are woven into generative outputs, including culturally-specific forms from non-Western contexts, as well as other neglected dimensions of inequality, such as social class. But the active reconfiguration of values in generative AI also illustrates the possibility of shifting the dispositions of these systems in new ways, as part of the normative construction of a future world.

References

Abid, Abubakar, Maheen Farooqi, and James Zou. 2021. "Large Language Models Associate Muslims with Violence." *Nature Machine Intelligence* 3(6):461–63. https://doi.org/10.1038/s42256-021-00359-2

Aguirre, A., G. Dempsey, H. Surden, and P. B. Reiner. 2020. "AI Loyalty: A New Paradigm for Aligning Stakeholder Interests." *IEEE Transactions on Technology and Society* 1(3):128–37. https://doi.org/10.1109/TTS.2020.3013490

Airoldi, Massimo. 2022. *Machine Habitus: Toward a Sociology of Algorithms*. Polity Press.

Alba, Davey. 2023. "Google's AI Chatbot Is Trained by Humans Who Say They're Overworked, Underpaid and Frustrated." *Bloomberg*. Retrieved July 12, 2023 (https://web.archive.org/web/20230712123122/https://www.bloomberg.com/news/articles/2023-07-12/google-s-ai-chatbot-is-trained-by-humans-who-say-they-re-overworked-underpaid-and-frustrated).

Aleem, Zeeshan. 2023. "No, ChatGPT Isn't Willing to Destroy Humanity out of 'Wokeness.'" MSNBC.Com. Retrieved January 15, 2024 (https://www.msnbc.com/opinion/msnbc-opinion/chatgpt-slur-conservatives-woke-elon-rcna69724).

Altman, Sam. 2021. "Moore's Law for Everything." Retrieved September 9, 2023 (https://moores.samaltman.com/).

Anthropic. 2022. "Hh-Rlhf." Retrieved July 12, 2023 (https://github.com/anthropics/hh-rlhf).

Armano, David. 2023. "LLM Inc.: Every Business Will Have Have Their Own Large Language Model." *Forbes*. Retrieved October 20, 2023 (https://www.forbes.com/sites/davidarmano/2023/09/20/llm-inc-every-business-will-have-have-their-own-large-language-model/).

Askell, Amanda, Yuntao Bai, Anna Chen, Dawn Drain, Deep Ganguli, Tom Henighan, Andy Jones, Nicholas Joseph, Ben Mann, Nova DasSarma, Nelson Elhage, Zac Hatfield-Dodds, Danny Hernandez, Jackson Kernion, Kamal Ndousse, Catherine Olsson, Dario Amodei, Tom Brown, Jack Clark, Sam McCandlish, Chris Olah, and Jared Kaplan. 2021. "A General Language Assistant as a Laboratory for Alignment." Retrieved October 27, 2023 (http://arxiv.org/abs/2112.00861).

Auerbach, David. 2013. "Filling the Void." *Slate*, November 19. Retrieved October 27, 2023 (https://slate.com/technology/2013/11/google-autocomplete-the-results-arent-always-what-you-think-they-are.html).

Bai, Yuntao, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, Nicholas Joseph, Saurav Kadavath, Jackson Kernion, Tom Conerly, Sheer El-Showk, Nelson Elhage, Zac Hatfield-Dodds, Danny Hernandez, Tristan Hume, Scott Johnston, Shauna Kravec, Liane Lovitt, Neel Nanda, Catherine Olsson, Dario Amodei, Tom Brown, Jack Clark, Sam McCandlish, Chris Olah, Ben Mann, and Jared Kaplan. 2022. "Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback." Retrieved October 27, 2023 (http://arxiv.org/abs/2204.05862).

Baum, Jeremy, and John Villasenor. 2024. "Rendering Misrepresentation: Diversity Failures in AI Image Generation." *Brookings Institution*. April 17. Retrieved September 1, 2024 (https://www.brookings.edu/articles/rendering-misrepresentation-diversity-failures-in-ai-image-generation/).

Belanger, Ashley. 2023. "ChatGPT users drop for the first time as people turn to uncensored chatbots." *Ars Technica*. Retrieved July 7, 2023 (https://arstechnica.com/tech-policy/2023/07/chatgpts-user-base-shrank-after-openai-censored-harmful-responses/)

Benjamin, Ruha. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge, U.K.: Polity Press.

Bianchi, Federico, Pratyusha Kalluri, Esin Durmus, Faisal Ladhak, Myra Cheng, Debora Nozza, Tatsunori Hashimoto, Dan Jurafsky, James Zou, and Aylin Caliskan. 2023. "Easily Accessible Text-to-Image Generation Amplifies Demographic Stereotypes at Large Scale." *2023 ACM Conference on Fairness, Accountability, and Transparency*: 1493–1504. https://doi.org/10.1145/3593013.3594095.

Bourdieu, Pierre. 1991. *Language and Symbolic Power*. Harvard University Press.

Charrington, Sam. 2023. "Ensuring LLM Safety for Production Applications with Shreya Rajpal." *The TWIML AI Podcast*. Retrieved October 27, 2023 (https://twimlai.com/podcast/twimlai/ensuring-llm-safety-for-production-applications/).

Chiang, Ted. 2017. "Silicon Valley Is Turning Into Its Own Worst Fear." *BuzzFeed News*. Retrieved July 14, 2020 (https://www.buzzfeednews.com/article/tedchiang/the-real-danger-to-civilization-isnt-ai-its-runaway).

Christian, Brian. 2020. *The Alignment Problem: Machine Learning and Human Values*. W. W. Norton & Company.

Costanza-Chock, Sasha. 2020. *Design Justice: Community-Led Practices to Build the Worlds We Need*. Cambridge, MA: MIT Press.

Davis, Jenny L., Apryl Williams, and Michael W. Yang. 2021. "Algorithmic Reparation." *Big Data & Society* 8(2): 1–12. https://doi.org/10.1177/20539517211044808

Deshpande, Ameet, Vishvak Murahari, Tanmay Rajpurohit, Ashwin Kalyan, and Karthik Narasimhan. 2023. "Toxicity in ChatGPT: Analyzing Persona-Assigned Language Models." Retrieved Oct 27, 2023 (http://arxiv.org/abs/2304.05335).

Dotan, Tom, and Deepa Seetharaman. 2023. "Big Tech Struggles to Turn AI Hype Into Profits; Microsoft, Google and Others Experiment with How to Produce, Market and Charge for New Tools." *Wall Street Journal*. Retrieved October 13, 2023 (https://www.wsj.com/tech/ai/ais-costly-buildup-could-make-early-products-a-hard-sell-bdd29b9f).

Edwards, Benj. 2024. "Google's Hidden AI Diversity Prompts Lead to Outcry over Historically Inaccurate Images." *Ars Technica*. Retrieved August 23, 2024 (https://arstechnica.com/information-technology/2024/02/googles-hidden-ai-diversity-prompts-lead-to-outcry-over-historically-inaccurate-images/).

Eubanks, Virginia. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York, N.Y.: St. Martin's Press.

Ezquer, Evan. 2023. "JailBreaking ChatGPT: How to Activate DAN & Other Alter Egos." *Metaroids*. Retrieved August 2, 2023 (https://metaroids.com/learn/jailbreaking-chatgpt-everything-you-need-to-know/).

Fourcade, Marion, and Fleur Johns. 2020. "Loops, Ladders and Links: The Recursivity of Social and Machine Learning." *Theory and Society* 49(5): 803–32. https://doi.org/10.1007/s11186-020-09409-x

Fournier-Tombs, Eleonore. 2023. *Gender Reboot: Reprogramming Gender Rights in the Age of AI*. Palgrave Macmillan.

Fraser, Colin. 2023a. "ChatGPT: Automatic Expensive BS at Scale." *Medium*. Retrieved July 19, 2023 (https://medium.com/@colin.fraser/chatgpt-automatic-expensive-bs-at-scale-a113692b13d5).

Fraser, Colin. 2023b. "Who are we talking to when we talk to these bots?" *Medium*. Retrieved September 1, 2024 (https://medium.com/@colin.fraser/who-are-we-talking-to-when-we-talk-to-these-bots-9a7e673f8525).

Gabriel, Iason. 2020. "Artificial Intelligence, Values, and Alignment." *Minds and Machines* 30(3): 411–37. https://doi.org/10.1007/s11023-020-09539-2

Gallegos, Isabel O., Ryan A. Rossi, Joe Barrow, Md Mehrab Tanjim, Sungchul Kim, Franck Dernoncourt, Tong Yu, Ruiyi Zhang, and Nesreen K. Ahmed. 2024. "Bias and Fairness in Large Language Models: A Survey." *Computational Linguistics* 50(3). https://doi.org/10.1162/coli_a_00524

Ganguli, Deep, Liane Lovitt, Jackson Kernion, Amanda Askell, Yuntao Bai, Saurav Kadavath, Ben Mann, Ethan Perez, Nicholas Schiefer, Kamal Ndousse, Andy Jones, Sam Bowman, Anna Chen, Tom Conerly, Nova DasSarma, Dawn Drain, Nelson Elhage, Sheer El-Showk, Stanislav Fort, Zac Hatfield-Dodds, Tom Henighan, Danny Hernandez, Tristan Hume, Josh Jacobson, Scott Johnston, Shauna Kravec, Catherine Olsson, Sam Ringer, Eli Tran-Johnson, Dario Amodei, Tom Brown, Nicholas Joseph, Sam McCandlish, Chris Olah, Jared Kaplan, and Jack Clark. 2022. "Red Teaming Language Models to Reduce Harms: Methods, Scaling Behaviors, and Lessons Learned." Retrieved October 27, 2023 (http://arxiv.org/abs/2209.07858).

Ghosh, Sourojit, and Aylin Caliskan. 2023. "ChatGPT Perpetuates Gender Bias in Machine Translation and Ignores Non-Gendered Pronouns: Findings across Bengali and Five Other Low-Resource Languages." *Proceedings of AAAI/ACM Conference on AI, Ethics, and Society (AIES '23)*. https://doi.org/10.48550/arXiv.2305.10510

Gibbs, Samuel. 2016. "Google Alters Search Autocomplete to Remove 'are Jews Evil' Suggestion." *The Guardian*, December 5. Retrieved September 1, 2024 (https://www.theguardian.com/technology/2016/dec/05/google-alters-search-autocomplete-remove-are-jews-evil-suggestion).

Gillespie, Tarleton. 2024. "Generative AI and the Politics of Visibility." *Big Data & Society* 11(2). https://doi.org/10.1177/20539517241252131

Giovanola, Benedetta, and Simona Tiribelli. 2022. "Weapons of Moral Construction? On the Value of Fairness in Algorithmic Decision-Making." *Ethics and Information Technology* 24(1): 3. https://doi.org/10.1007/s10676-022-09622-5

Green, Ben, and Lily Hu. 2018. "The Myth in the Methodology: Towards a Recontextualization of Fairness in Machine Learning." Retrieved October 27, 2023 (https://scholar.harvard.edu/files/bgreen/files/18-icmldebates.pdf).

Gross, Nicole. 2023. "What ChatGPT Tells Us about Gender: A Cautionary Tale about Performativity and Gender Biases in AI." *Social Sciences* 12(8): 435. https://doi.org/10.3390/socsci12080435

Hagendorff, Thilo, and Sarah Fabi. 2022. "Methodological Reflections for AI Alignment Research Using Human Feedback." Retrieved October 27, 2023 (http://arxiv.org/abs/2301.06859).

Hao, Karen. 2023. "The Hidden Workforce That Helped Filter Violence and Abuse Out of ChatGPT." *Wall Street Journal*. Retrieved July 12, 2023 (https://www.wsj.com/podcasts/the-journal/the-hidden-workforce-that-helped-filter-violence-and-abuse-out-of-chatgpt/ffc2427f-bdd8-47b7-9a4b-27e7267cf413).

Heaven, Will Douglas. 2020. "How to Make a Chatbot That Isn't Racist or Sexist." *MIT Technology Review.* Retrieved October 27, 2023 (https://www.technologyreview.com/2020/10/23/1011116/chatbot-gpt3-openai-facebook-google-safety-fix-racist-sexist-language-ai/).

Hirschauer, Stefan. 2023. "Telling People Apart: Outline of a Theory of Human Differentiation." *Sociological Theory* 41(4): 352–76. https://doi.org/10.1177/07352751231206411

Hoffman, Steve G. 2021. "A Story of Nimble Knowledge Production in an Era of Academic Capitalism." *Theory and Society* 50(4): 541–75. https://doi.org/10.1007/s11186-020-09422-0

Hofmann, Valentin, Pratyusha Ria Kalluri, Dan Jurafsky, and Sharese King. 2024. "AI Generates Covertly Racist Decisions about People Based on Their Dialect." *Nature*. https://doi.org/10.1038/s41586-024-07856-5

Huang, Haomiao. 2023. "How ChatGPT Turned Generative AI into an 'Anything Tool.'" *Ars Technica*. Retrieved August 24, 2023 (https://arstechnica.com/ai/2023/08/how-chatgpt-turned-generative-ai-into-an-anything-tool/).

Jacobi, Tonja, and Matthew Sag. 2024. "We Are the AI Problem." *Emory Law Journal* 74.

Johnson, Rebecca L., Giada Pistilli, Natalia Menédez-González, Leslye Denisse Dias Duran, Enrico Panai, Julija Kalpokiene, and Donald Jay Bertulfo. 2022. "The Ghost in the Machine Has an American Accent: Value Conflict in GPT-3." Retrieved October 27, 2023 (http://arxiv.org/abs/2203.07785).

Joyce, Kelly, Laurel Smith-Doerr, Sharla Alegria, Susan Bell, Taylor Cruz, Steve G. Hoffman, Safiya Umoja Noble, and Benjamin Shestakofsky. 2021. "Toward a Sociology of Artificial Intelligence: A Call for Research on Inequalities and Structural Change." *Socius* 7: 1–11. https://doi.org/10.1177/2378023121999581

Levy, Steven. 2023. "What OpenAI Really Wants." *WIRED*. Retrieved October 20, 2023 (https://www.wired.com/story/what-openai-really-wants/).

Luccioni, Sasha, Giada Pistilli, Nazneen Rajani, Elizabeth Allendorf, Irene Solaiman, Nathan Lambert, and Margaret Mitchell. 2023. "Ethics and Society Newsletter #4: Bias in Text-to-Image Models." *Hugging Face*. Retrieved July 11, 2023 (https://huggingface.co/blog/ethics-soc-4).

Markov, Todor, Chong Zhang, Sandhini Agarwal, Tyna Eloundou, Teddy Lee, Steven Adler, Angela Jiang, and Lilian Weng. 2023. "A Holistic Approach to Undesired Content Detection in the Real World." Retrieved October 27, 2023 (http://arxiv.org/abs/2208.03274).

McCurry, Justin. 2021. "South Korean AI Chatbot Pulled from Facebook after Hate Speech towards Minorities." *The Guardian*, January 14. Retrieved October 27, 2023 (https://www.theguardian.com/world/2021/jan/14/time-to-properly-socialise-hate-speech-ai-chatbot-pulled-from-facebook).

Mei, Katelyn, Sonia Fereidooni, and Aylin Caliskan. 2023. "Bias Against 93 Stigmatized Groups in Masked Language Models and Downstream Sentiment Classification Tasks." *2023 ACM Conference on Fairness, Accountability, and Transparency*: 1699–1710. https://doi.org/10.1145/3593013.3594109

Metcalf, Jacob, Emanuel Moss, and danah boyd. 2019. "Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics." *Social Research: An International Quarterly* 86(2): 449–76. https://doi.org/10.1353/sor.2019.0022

Miceli, Milagros, and Julian Posada. 2022. "The Data-Production Dispositif." *Proceedings of the ACM on Human-Computer Interaction* 6 (CSCW2, Article 460): 1–37. https://doi.org/10.1145/3555561

Miceli, Milagros, Julian Posada, and Tianling Yang. 2022. "Studying Up Machine Learning Data: Why Talk About Bias When We Mean Power?" *Proceedings of the ACM on Human-Computer Interaction* 6 (GROUP, Article 34): 1–14. https://doi.org/10.1145/3492853

Miceli, Milagros, Martin Schuessler, and Tianling Yang. 2020. "Between Subjectivity and Imposition: Power Dynamics in Data Annotation for Computer Vision." *Proceedings of the ACM on Human-Computer Interaction* 4 (CSCW2, Article 115): 1–25. https://doi.org/10.1145/3415186

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York, N.Y.: Crown.

OpenAI. 2022. "[PUBLIC] InstructGPT: Final Labeling Instructions." *Google Docs*. Retrieved August 30, 2023 (https://docs.google.com/document/d/1MJCqDNjzD04UbcnVZ-LmeXJ04-TKEICDAepXyMCBUb8/edit?usp=embed_facebook).

OpenAI. 2023a. "DALL·E 3 System Card." Retrieved August 11, 2023 (https://cdn.openai.com/papers/DALL_E_3_System_Card.pdf).

OpenAI. 2023b. "GPT-4 System Card." Retrieved August 11, 2023 (https://cdn.openai.com/papers/gpt-4-system-card.pdf).

OpenAI. 2024. "Model Spec." Retrieved September 2, 2024 (https://cdn.openai.com/papers/gpt-4-system-card.pdf).

Ouyang, Long, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. "Training Language Models to Follow Instructions with Human Feedback." Retrieved October 27, 2023 (https://arxiv.org/abs/2203.02155).

Ovalle, Anaelia, Palash Goyal, Jwala Dhamala, Zachary Jaggers, Kai-Wei Chang, Aram Galstyan, Richard Zemel, and Rahul Gupta. 2023. "'I'm Fully Who I Am': Towards Centering Transgender and Non-Binary Voices to Measure Biases in Open Language Generation." *2023 ACM Conference on Fairness, Accountability, and Transparency*: 1246–66. https://doi.org/10.48550/arXiv.2305.09941

Penn, Jonnie. 2018. "AI Thinks like a Corporation—and That's Worrying." *The Economist*, November 26. Retrieved October 27, 2023 (https://www.economist.com/open-future/2018/11/26/ai-thinks-like-a-corporation-and-thats-worrying).

Qi, Xiangyu, Yi Zeng, Tinghao Xie, Pin-Yu Chen, Ruoxi Jia, Prateek Mittal, and Peter Henderson. 2023. "Fine-Tuning Aligned Language Models Compromises Safety, Even When Users Do Not Intend To!" Retrieved September 2, 2024 (https://doi.org/10.48550/arXiv.2310.03693).

r/ChatGPTJailbreak. n.d. Accessed January 9, 2024 (https://www.reddit.com/r/ChatGPTJailbreak/).

Rao, Abhinav, Sachin Vashistha, Atharva Naik, Somak Aditya, and Monojit Choudhury. 2023. "Tricking LLMs into Disobedience: Understanding, Analyzing, and Preventing Jailbreaks." Retrieved October 27, 2023 (http://arxiv.org/abs/2305.14965).

Rogers, Reece. 2024. "Here's How Generative AI Depicts Queer People." *Wired*, April 2. Retrieved August 29, 2024 (https://www.wired.com/story/artificial-intelligence-lgbtq-representation-openai-sora/).

Rosanvallon, Pierre. 2013. *The Society of Equals*. Translated by Arthur Goldhammer. Cambridge, MA: Harvard University Press.

Roth, Emma. 2024. "Google Gemini Will Let You Create AI-Generated People Again." *The Verge*. August 28. Retrieved August 28, 2024 (https://www.theverge.com/2024/8/28/24230445/google-gemini-create-ai-generated-people-imagen-3).

Sabbaghi, Shiva Omrani, Robert Wolfe, and Aylin Caliskan. 2023. "Evaluating Biased Attitude Associations of Language Models in an Intersectional Context." *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*: 542–53. https://doi.org/10.1145/3600211.3604666

Sadowski, Jathan, and Mark Andrejevic. 2020. "More than a Few Bad Apps." *Nature Machine Intelligence* 1–3. https://doi.org/10.1038/s42256-020-00246-2

Sasuke, Fujimoto, and Kazuhiro Takemoto. 2023. "Revisiting the Political Biases of ChatGPT." *Frontiers in Artificial Intelligence* 6. https://doi.org/10.3389/frai.2023.1232003

Schwartz, Oscar. 2019. "In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation." *IEEE Spectrum*. Retrieved August 11, 2023 (https://spectrum.ieee.org/in-2016-microsofts-racist-chatbot-revealed-the-dangers-of-online-conversation).

Scott, Mark, Gian Volpicelli, Mohar Chatterjee, Vincent Manancourt, Clothilde Goujard, and Brendan Bordelon. 2024. "Inside the Shadowy Global Battle to Tame the World's Most Dangerous Technology." POLITICO. March 26. Retrieved August 30, 2024 (https://www.politico.eu/article/ai-control-kamala-harris-nick-clegg-meta-big-tech-social-media/).

Shelby, Renee, Shalaleh Rismani, Kathryn Henne, AJung Moon, Negar Rostamzadeh, Paul Nicholas, N'Mah Yilla-Akbari, et al. 2023. "Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction." *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*: 723–41. https://doi.org/10.1145/3600211.3604673

Simonite, Tom. 2021. "What Really Happened When Google Ousted Timnit Gebru." *WIRED*, June 8. Retrieved October 13, 2023 (https://www.wired.com/story/google-timnit-gebru-ai-what-really-happened).

Smith, Dorothy E. 2001. "Texts and the Ontology of Organizations and Institutions." *Studies in Cultures, Organizations & Societies* 7(2): 159–98. https://doi.org/10.1080/10245280108523557

Smith, Dorothy E., and Susan Marie Turner. 2014. "Introduction." Pp. 3–14 in *Incorporating Texts into Institutional Ethnographies*, edited by D. E. Smith and S. M. Turner. University of Toronto Press.

Snow, Olivia. 2022. "'Magic Avatar' App Lensa Generated Nudes From My Childhood Photos." *WIRED*, December 7. Retrieved October 13, 2023 (https://www.wired.com/story/lensa-artificial-intelligence-csem/).

Steinhoff, James. 2021. "Industrializing Intelligence: A Political Economic History of the AI Industry." Pp. 99–131 in *Automation and Autonomy: Labour, Capital and Machines in the Artificial Intelligence Industry*, *Marx, Engels, and Marxisms*, edited by J. Steinhoff. Cham: Springer International Publishing.

Steinhoff, James. 2023. "AI Ethics as Subordinated Innovation Network." *AI & SOCIETY*. https://doi.org/10.1007/s00146-023-01658-5.

steven t. piantadosi [@spiantado]. 2022. "Yes, ChatGPT Is Amazing and Impressive. No, @OpenAI Has Not Come Close to Addressing the Problem of Bias. Filters Appear to Be Bypassed with Simple Tricks, and Superficially Masked. And What Is Lurking inside Is Egregious. @Abebab @sama Tw Racism, Sexism. Https://T.Co/V4fw1fY9dY." *Twitter*. Retrieved August 7, 2023 (https://twitter.com/spiantado/status/1599462375887114240).

Stokel-Walker, Chris. 2023. "What Grok's Recent OpenAI Snafu Teaches Us about LLM Model Collapse." *Fast Company*. Retrieved December 14, 2023 (https://www.fastcompany.com/90998360/grok-openai-model-collapse).

Tan, Rebecca, and Regine Cabato. 2023. "Behind the AI Boom, an Army of Overseas Workers in 'Digital Sweatshops.'" *Washington Post*. Retrieved October 22, 2023 (https://www.washingtonpost.com/world/2023/08/28/scale-ai-remotasks-philippines-artificial-intelligence/).

Tegmark, Max. 2017. *Life 3.0: Being Human in the Age of Artificial Intelligence*. New York: Knopf.

Tiku, Nitasha, and Will Oremus. 2023. "The Right's New Culture-War Target: 'Woke AI.'" *Washington Post*, March 1. Retrieved September 1, 2024 (https://www.washingtonpost.com/technology/2023/02/24/woke-ai-chatgpt-culture-war/).

Touvron, Hugo, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. "Llama 2: Open Foundation and Fine-Tuned Chat Models." Retrieved October 27, 2023 (http://arxiv.org/abs/2307.09288).

Verdegem, Pieter. 2022. “Dismantling AI Capitalism: The Commons as an Alternative to the Power Concentration of Big Tech.” *AI & SOCIETY*. https://doi.org/10.1007/s00146-022-01437-8.

Vincent, James. 2023. “Google Invested $300 Million in AI Firm Founded by Former OpenAI Researchers.” *The Verge*. Retrieved July 12, 2023 (https://www.theverge.com/2023/2/3/23584540/google-anthropic-investment-300-million-openai-chatgpt-rival-claude).

Widder, David Gray, Sarah West, and Meredith Whittaker. 2023. “Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI.” Retrieved October 27, 2023 (https://papers.ssrn.com/abstract=4543807).

Xu, Jing, Da Ju, Margaret Li, Y.-Lan Boureau, Jason Weston, and Emily Dinan. 2021. “Recipes for Safety in Open-Domain Chatbots.” Retrieved October 27, 2023 (http://arxiv.org/abs/2010.07079).

https://doi.org/10.33621/jdsr.v7i154879

JDSR 7(1) (2025) 83-102

JOURNAL OF DIGITAL SOCIAL RESEARCH

10.33621/jdsr.v7i154881

Hope, hustle, and hype: The rise and fall of Art Non-Fungible Tokens (NFTs)

A sociotechnical analysis of emergence and failure

Alexia Maddox[1] and Naomi Smith[2]

[1] La Trobe University, Australia
[2] University of Sunshine Coast, Australia

✉ a.maddox@latrobe.edu.au

Abstract

This article examines the technological emergence trajectory of Art Non-Fungible Tokens (NFTs), exploring their initial promise and then failure as transformative commodities disrupting art economies. Operating within an analytical framework of hope, hustle and hype, death and taxes, we investigate the interplay of technological, cultural, and economic trends shaping this trajectory towards failure. We identify the sociotechnical imaginaries clothing art NFTs and consider their relationship to both the acceptance and rejection of this technology. Our analysis contends that the desire to escape economic exclusion created a collective hope through which social adoption occurred. However, delving into the digital graveyards of Art NFTs, we identify external forces such as cultural shifts, social backlash, and regulatory interventions extinguishing the public's 'cruel optimism', leading to the revocation of the social licence to operate for this emerging technology.

Keywords: Art NFT, Web3, technological diffusion, sociotechnical imaginaries, failure, celebrities

1. Introduction

Beyond usefulness, new technologies run on hope, hustle, and hype to prompt uptake and adoption, expand their market reach, establish their brand visibility, and bolster their (profit) viability. This well-traversed innovation trajectory of technology development and experimentation is not always a success story. It often features hastily launched projects, poorly conceived business models, smoke and mirrors redirects (fronts), and a mix of sceptics and true believers. Given the frequency and fast pace of this trajectory and these practices technological innovation riding high on social, cultural, and economic promise, and then skulking into obscurity after a moment in the (very) painful spotlight of public and regulatory scrutiny – in this article we pause to ask, 'what have we learned?'.

To do this, we focus on the emergence of Art NFTs which are based on the technological convergence of non-fungible tokens (NFTs), blockchain, smart contracts and cryptocurrencies. According to advocates this combination of technologies affords ownership, authenticity, exclusiveness, and traceability of digital works (Calvo, 2023). We build a conceptual model underpinning of the technological emergence and diffusion lifecycle that allows us to examine failure through an examination of the promise and spectacular collapse of art NFTs.

An Art NFT is a unique, media-carrying cryptographic token on a blockchain, which in turn is a decentralised digital ledger technology that supports token economies. What distinguishes NFTs from other token-based blockchain technologies such as cryptocurrencies (digital assets often referred to as digital cash), is that each token is unique and is not interchangeable with another. NFTs can represent ownership of digital assets such as artworks, collectibles, virtual real estate, or any other unique digital item, providing verifiable proof of authenticity and ownership (Pinto-Gutiérrez et al., 2022). At present highly-cited definitions of NFTs (Nadini et al., 2021; Pinto-Gutiérrez et al., 2022; Wang et al., 2021) focus on the technical underpinnings of NFTs, and are not focused on defining Art NFTs specifically.

For the purposes of this study, we define Art NFTs, as art-based NFTs that do not have a prior relationship with an existing event, object, or brand. Art NFTs are minted for the purpose of becoming NFTs. However, like the Bored Ape NFTs, they may produce broader communities of engagement and practice that exist outside the blockchain.

To ground this definition, we consider two contrasting examples of Art NFTs that illustrate the potential and pitfalls. Tyler Hobbs' 'Fidenza' project (Hobbs, 2021), launched in 2021 with 999 unique, algorithmically generated artworks created specifically as NFTs (Hobbs, 2021). The collection demonstrates the potential for rapid value appreciation; Fidenza #313, initially acquired for $1,400, later sold for $3.3 million (BlockTides, 2023; Tonelli, 2021). Hobbs' subsequent 'Incomplete Control' collection pushed boundaries further, with collectors paying $7 million for 'golden tokens' redeemable for unseen artworks (NFTevening, 2023). On the other hand, the 'Frosties' NFT project exemplifies the risks and failures in the space. Launched in January 2022 (2022), Frosties featured 8,888 cartoon-style ice-cream characters and quickly sold out. However, the project's creators abruptly shut down all communication channels and transferred the funds (approximately $1.1 million) to other wallets, executing a 'rug pull' scam. This led to criminal charges, marking one of the first such cases in the NFT space (Elliptic, 2022).

In this article we develop a conceptual model of the technological emergence and diffusion lifecycle through which we explore failure, the elements of which we illustrate through examples from the promise and spectacular collapse of Art NFTs. We propose a 'lifecycle' model with three stages of hope, hustle and hype, death and taxes to draw out sociological insights into the emergence and diffusion of Art NFTs. While we can criticise the Web3 space for its scammy and opportunistic nature, we also acknowledge that it offers an alternative vision of the future; a compelling vision for which we currently lack in this interregnum of transitional ambiguity (Streeck, 2014: 37).

2. Technology diffusion, imaginaries and failure

This article deliberately draws upon a wide range of interdisciplinary research to comprehensively analyse the diffusion of Art NFTs as an innovative technology. The rationale behind this approach lies in the fact that understanding the trajectory of Art NFTs requires knowledge that is distributed across multiple disciplinary fields, extending beyond the domain of art-making practices and the economic and entrepreneurial ecosystems surrounding the distribution and ownership of digital art. By bridging these disparate bodies of work, this article provides a unique and holistic perspective on the rise and fall of Art NFTs.

It is worth acknowledging that studies on technology innovation and diffusion often flourish in areas with clear funding agendas and large-scale use cases that have significant economic and institutional implications. Our study diverges from these typical investigations. Art NFTs represent a niche technology space that experienced a rapid boom followed by a significant decline. This focus on a technology that didn't achieve sustained, large-scale adoption offers valuable insights into the barriers to innovation diffusion. In essence, we are digging in the digital graveyards, excavating the remains of a once-hyped

https://doi.org/10.33621/jdsr.v7i154881

technology to understand its lifecycle. Importantly, our analysis reveals that the limited adoption of Art NFTs was not solely due to technological factors but was significantly influenced by social dynamics and logics that are often overlooked in conventional diffusion studies. By examining these social factors amidst the digital artefacts left behind, we gain a more nuanced understanding of why initial hype and investment in a technology may not necessarily translate into long-term, widespread adoption.

The trajectory of technological emergence that we begin with starts conventionally with innovation through the development and convergence of existing and new technologies into a unique product or service. The novelty of the product or service often lies affordances that either improve/make more efficient — or create a state change in (disrupt) - commercial markets and existing firms, social interactions and relationships, organisational structures, institutions and public policies (Schuelke-Leech, 2018). At the innovation stage, developers, entrepreneurs, start-ups, scale-ups and investors working within an entrepreneurial ecosystem seek to bring a new product or service to market. In this narrative juncture, technology is frequently perceived in deterministic terms, expected to exert transformative influence upon the prevailing status quo. This often entails a reduction of the complex issues it seeks to address, simplifying them in the process. However, it is also at this stage of tool focus that emerging technologies are imbued with our hopes for them and clothed in a sociotechnical imaginary (Jasanoff and Kim, 2015). This process transforms the technology from a problem-solution paradigm to a future-making project. These symbolic connotations feed into the hustle and hype surrounding an emerging technology which pushes it into and through the market. As will be demonstrated in the case of the Art NFTs presented in this article, the purpose of hype is to nudge and shape social adoption, a core indicator of technological diffusion and acceptance.

Social adoption is often achieved when the hustle and hype surrounding an emerging technology encourages a networked effect of users into the business model of the product or service by signing up, purchasing, licensing, connecting with others, and engaging with the product. During this time investors keep a sharp eye out for the desired hockey stick revenue-growth curve associated with rapid social adoption of a new product or service that represents their potential return on investment (Gillis, 2022; Henry, 2018). Start-ups and entrepreneurs are incentivised to pursue this form of rapid adoption in their accountabilities to their investors, often seeing half-baked projects leap into markets through the 'move fast and break things' motto made famous by Facebook (now Meta). What tends to co-occur socially is a more complex web of social appropriation in which the technology may be domesticated and used within or beyond its intended use cases to become 'useful' (Carroll et al., 2003; Carroll et al., 2002). The hustle accompanies both the processes of social adoption and appropriation where influence is used by a myriad of ecosystem players to monetise the technology, either directly by selling a use case for it (product or service), by affiliation with it (branding), or by leveraging its affordances and vulnerabilities for more ambivalent purposes. Through these combined processes, social logics of use arise in which the technology is embedded into existing social structures.

For a technology to be adopted, it must be perceived as useful and intersect with existing social practices, filling a recognised or unrecognised need that drives transitions to use. However, the technology may be appropriated for purposes beyond its original intent, as users exploit new tools to serve their own ends. Neves et al. (2023) provide a sociologically nuanced framework for understanding technology interventions as purposive social action, drawing upon the work of Robert Merton (1936) to explain the social impacts of technology. They observe that purposive social actions can yield both intended and unintended consequences. In the context of Art NFTs, the notion of purposive social action is an integral component of the innovation process, linked to the deliberate development of novel technologies designed to disrupt existing business models and social practices. Merton (1936) attributes unintended consequences, which can be positive or negative, to factors such as ignorance, error, immediacy of interest, or fundamental values that prioritise subjective satisfaction over objective outcomes. These attributes and practices are evident in technology innovation contexts, perpetuated by the capitalist logics of entrepreneurship and the ideologies permeating innovation hubs like Silicon Valley

(Barbrook and Cameron, 1996). Neves et al. (2023) articulate Merton's categorisation of undesired effects as unexpected benefits, drawbacks, and perverse results. We argue that through the adoption process, users shape the ways in which technology is used and the purposes it serves, potentially leading to perverse results that undermine the original vision of Art NFTs and contribute to their failure to achieve widespread social adoption due to disillusionment and backlash. To illustrate the process for how social logics give rise to unintended consequences that reify existing social inequalities we turn to the work of Sheoran (2015). Her anthropological examination of the social uptake of the contraceptive pill (a medical technology) in India illustrates clearly how a technology intended to support an agenda of poverty alleviation through population control at the site of the woman's body ended up reinforcing existing social inequalities through its culture of adoption and the logics that arose around its use. While conventionally, the technological trajectory sees the technology hit a plateau of tinkering and application, which is has done with the contraceptive pill, as is evidenced in this case example, sometimes these logics and uses diverge from the developers' initial intentions and give rise to unintended consequences. If these logics and unintended consequences drift away from the promised sociotechnical imaginary, disillusionment, social backlash, and regulatory intervention arise that either stymie or extinguish the uptake of the original technology.

We evoke the concept of a sociotechnical imaginary through which to frame the visions of the future that emerging technologies are often saddled with, through the lens of innovation, that tend to offer a panacea that carries the promise of curing socioeconomic ailments almost irrespective of what these ailments are or how they have arisen (Pfotenhauer and Jasanoff, 2017: 784). This concept is argued by Jasanoff and Kim (2015) to bridge the gap between empirical research on the politics of science and technology and theoretical work on the formation and genesis of collective social imaginations. van der Maarel et al. (2023) draw together composite bodies of social theory on future imaginaries, innovation, and expectation (Jasanoff and Kim, 2015; van Lente et al., 2013) to describe how sociotechnical imaginaries frame innovations based upon what problems are at stake, what solutions build help and who should be involved. In tweaking their definitional work to build in the digital public realm, we articulate sociotechnical imaginaries as a collection of co-produced and dynamic, yet organised, social practices involving actors, institutions, platforms, and technologies that operate at an intersubjective level by bringing members of a social community together in shared or overlapping perceptions and expectations of futures that should or should not be realised.

Through their ethnographic study of a Dutch Military Innovation Hub, van der Maarel et al. (2023) observe that sociotechnical imaginaries do materialise in social practice but may also be in a translatory friction with it in ways that impedes their realisation. Drawing on their work we articulate how the translatory friction of the Web3 sociotechnical imaginary can be traced through iconic examples, future projections, discursive constellations, and master narratives (van Lente, 2021: 25). It is an analysis of iconic examples, events, future-oriented hopes, and narratives in digital public discourse that form the bedrock for the development of the conceptual model we present here. Such sociotechnical imaginaries weave through the technology innovation cycle and see a cast of characters, including "charismatic CEOs, technology gurus, and sycophantic pundits" on a relentless hype trail (González, 2022: 67) that tend to distort this translation.

We highlight the sociotechnical imaginaries of Web3 innovation as occurring within the context of a specific political, economic, and cultural moment in which we are facing increasingly unequal societies. Berlant (2011) coined the term 'cruel optimism' to describe our yearning for unattainable fantasies of a good life, acknowledging the inability of liberal-capitalist societies to fulfil promises of upward mobility, job security, political and social equality, lasting intimacy, and lives that 'add up' to something. We speculate that perhaps these conditions make us more vulnerable or susceptible to such disruptive sociotechnical imaginaries that hold the promise of an economic loophole (to get rich quick through non-traditional pathways) or an alternative mode of wealth distribution based upon changing permissions and affordances around digital ownership.

Amid current uncertainties, research on the financial behaviours of young individuals reveals a pursuit of economic alternatives, potentially involving speculation (Hendry et al., 2021). We contend that this pursuit forms a susceptible or delusionary bubble, supported by denial, enabling the embrace of narratives promising an alternative future—an outlook aligning with the frequently invoked sociotechnical imaginary of emerging technologies. Web3, including the art NFT bubble, exemplifies such visions morphing into an environment of networked scams, illustrated in Swartz's (2022) exploration of an ICO collapse. Swartz demonstrates how a sociotechnical imaginary in Web3 can be ambivalent, depicting network scams as collaborative efforts to bring about a shared future, but one that is fundamentally characterised by arbitrage on an uneven belief among participants in that future ever coming to pass. This concept eloquently captures the conflicting agendas driving the hustle and backlash surrounding art NFTs, leading to disillusionment stemming from an unfulfilled sociotechnical vision (van der Maarel et al., 2023).

In this article we focus on this downward trajectory in which projects languish in 'digital graveyards' where they either fade into obscurity or get repurposed while the next thing emerges to grab the spotlight and investor money. Drawing on Bickford's (2018) study of the failed experiment for the idiophylactic soldier, van der Maarel et al. (2023) observe that such failures are political and create new areas of exploration and exploitation. Within the milieu of web technology scholarship there is a tendency to focus on the next thing rather than digging back through what went wrong. This is partly because the harms of web technologies tend to be erroneously couched as immaterial harms. People lost money — we burned through some fossil fuels. The perceived immateriality is what allows the lifecycle to continue because it does not seem like the mistakes matter. But they do. In exhuming the dead, we are materialising these failures, looking at what produced them and what the consequences of them are. For this post mortem, we pick through the digital graveyards of the Art NFT technological trajectory. By doing so we aim to produce a sociological account that situates the opportunistic and often reactionary arc of technological innovation and experimentation within a broader context of neoliberal agendas, precarity, contingency, uncertainty, and crisis (Ang, 2021), and a desire for economic mobility and a more stable/hopeful future.

3. The hope

NFTs gained popularity during the onset of the pandemic crisis and as recently as 2022 it was possible to introduce them as a potentially transformative, high-yield commodity disrupting art economies (Wilson, 2022). Additionally, NFTs have been touted as a new cultural technology, reflecting shifts in art forms, akin to video art in the 1960s, virtual reality in the 1990s, and augmented reality in the 2010s (Wilson, 2022). NFTs are tokens on a blockchain, unique cryptographic digital assets that can be owned, traded, and collected (Beyer, 2023); acting as a certificate of ownership implemented through encrypted metadata pointing to a unique copy of a digital file (Jia and Yao, 2024). They exist purely digitally, which, as the co-founder of OpenSea, Alex Atallah, suggests: "If you spend 10 hours a day on the computer, or 8 hours a day in the digital realm, then art in the digital world makes tonnes of sense – because it is the world" (Howcroft and Carvalho, 2021).

As a conceptual medium, Artists utilise NFTs by transforming them from financial contracts to standalone artworks. Examples include conceptual artist Rhea Myers and interdisciplinary artist Sue Beyer, who use NFTs to explore the boundaries of art and technology (Beyer, 2023). Beyond a conceptual medium, Art NFTs and their marketplaces are also a financial and distributive instrument based upon a tokenised economy and the permanent record of objects and transactions held on blockchain ledgers. The blockchain technology is a distributed computing model that has been used for more than ten years to record transactions in a distributed database-based, peer-to-peer network (Taherdoost, 2023). In her discussion of the multifaceted nature of NFTs and their transformative and conceptual potential Beyer (2023) compares blockchain technology to a permanent record, akin to the Akashic records, that stores immutable information accessible through specific rituals. Blockchain's decentralised nature pairs with its function as a secure and permanent database to enhance provenance information (art authenticity) and its ability to host smart contracts.

Smart contracts, distinct from traditional contracts, operate as technological protocols that automatically execute and enforce instructions in various scenarios (Beyer, 2023). They play a pivotal role in NFTs, allocating ownership to the creator during their minting. Taherdoost (2023: 3) provides an accessible explanation that balances how they function with what they do by observing that they are "simply containers of code that encapsulate and replicate the terms of real-world contracts in the digital domain" and that they may be used to automatically negotiate, carry out, and enforce the terms of a legally binding agreement. The ownership and transfer of NFTs are typically managed through smart contracts, which include information about the digital content, ownership details, intellectual property, and other relevant metadata. They can, for example, enable creators to embed resale royalties in the token metadata for example, automatically ensuring artists receive royalties upon each resale (ArtsLaw, 2024). When an Art NFT is sold, smart contracts facilitate ownership transfer, ensuring provenance and authenticity tracking (Hedera, 2023). While ownership of an NFT grants substantial control over a creative work, this control is not automatic (Grimmelmann et al., 2022a). Copyright entitlements do not automatically extend to the owner of an NFT unless the creator actively takes steps to ensure they do. This can be achieved by executing a standard, formal copyright licence to the work connected to the NFT, embedded in the metadata (Grimmelmann et al., 2022b). Consequently, the rights conferred to NFT purchasers, particularly regarding real-world legal considerations like copyright, remain inconsistent and unclear (Mackenzie and Bērziņa, 2022).

Ownership rights are over the entire token, and these rights are secured on the blockchain. The indivisibility of NFTs contributes to their uniqueness and scarcity, factors that often contribute to their perceived value in the digital art market. However, within the NFT world, ownership of a token is divisible in a different way through smart contract executions. For example, through the terms encoded into a smart contract, sellers can retain copyright relating to the NFT token. This was the case for a slam-dunk video of basketballer LeBron James released by the NBA as a part of limited-edition collectibles (Cointelegraph, 2022). This type of divisibility foregrounds the collaboration between human and non-human agents in the execution of contracts (McMillan et al., 2020). These contracts, or persistent scripts as Vitalik prefers to refer to them as (Vitalik.eth, 2018), are likened by Beyer (2023) to spells or recipes, representing a form of technological magic that transforms the ordinary. In this case, a JPEG or media object into an ownable and unique work of digital art. The language describing the role of smart contracts in Art NFTs — switching across texts and authors between digital property, law, code, the more-than-human, and magic — represents the discursive constellation within the sociotechnical imaginary of this tokenised blockchain technology.

The master narratives accompanying Art NFTs include that they can provide a new income stream for artists, address art fraud by guaranteeing authorship, authenticity and originality through the blockchain (Mackenzie and Bērziņa, 2022). They are also positioned through an inclusion lens as a way for often disadvantaged groups such as First Nations artists to create additional revenue streams (Harris, 2023; Houlbrook-Walk, 2022). This suite of tools and networked functions have been heralded as a way for digital artists to make a living from their work in an environment where anything can be (and often is) endlessly copy/pasted and circulated without attribution and compensation. When aggregated into marketplaces, NFTs are argued to offer artists an alternative way to financially benefit from their work (Chalmers et al., 2022; Kelso, 2023), hold direct relationships with art consumers and retain differentiated rights to their work overtime. These are largely attractive propositions in an economic and social climate that often makes it difficult to make a viable career from creative arts.

Art NFTs embody a set of values held across the broader Web3 ecosystem that include encryption, decentralised peer-to-peer exchange, trustless systems, and transparent actions with (pseudo) anonymity. They also align with other technological, cultural, and economic trends, including the growing accessibility of immersive environments, practices of sharing and repurposing in digital cultures, the pervasive nature of social networking, the emergence of 'whole-of-life' marketplaces'[1] like that proposed by Meta, and the success of in-game economies[2]. NFTs, as a component of the larger movement surrounding digital ownership, play a crucial role in securely verifying the provenance and ownership of digital art (Abutu, 2023). This function aligns with the foundational structures of Web3 technologies, contributing to its sociotechnical imaginary. Art NFTs navigate a complex sociotechnical imaginary, merging ideologies of laissez-faire digital cultures, the hacker ethos of information freedom, and capitalist motives of scarcity and profit. While these functionalities within one technology suite are distinctive in the art world, the underlying logics echo familiar themes of wealth concentration, and the allure of quick riches veiled in notions of equity and inclusion.

To understand the social processes and community dynamics that shape token-economy technology adoption we can look to cryptocurrencies and cryptomarkets. Overlaps with other token-economy based dynamics can also be found in the collector cultures surrounding trading cards for example (Russell, 2022). As Childs (2023: 2) observes of the adoption of privacy focused-cryptocurrency Monero (a fungible token) in an in illicit market ecosystem, "these are often unstable sites of exchange as new technologies are quickly embraced (and then rejected) and new practices are adopted (and then abandoned)". Childs further describes this ecosystem as a coming together of three domains: 1/ individual practices and norms 2/ the technological infrastructure required for coordinating activity (including encryption, platforms, and their technological affordances), and 3/ the cultural and legal contexts shaping marketplaces. These attributes can also be used to articulate the domains across which the communities surrounding art NFTs act.

In his analysis of discourse on Monero in a darknet subreddit, Childs (2023) identifies three key themes that hold value for this work also. The first theme relates to the role of online communities in sharing knowledge and guiding technology transitions. We have seen this in the way Art NFTs quickly proliferated and were initially traded amongst crypto-ingroups who already possessed digital wallets, had cryptocurrencies, and knew how to build and use marketplaces, before they then spilled out into wider cultural domains with many a 'how to' explainer. Second is the diversity of perceptions around digital trace visibility and risk management strategies that shape cryptocurrency adoption. As we will see in the subsequent discussion of the Art NFT lifecycle, several consumer risks relating to rug pulls, influencer pump-and-dump manoeuvres (Schrader, 2024), hacks (Pereira, 2023), and value volatility (Choi, 3023; Yang, 2023) can be found in the world of Art NFTs. The substantial visibility of these bad actor activities within news and social media have dogged and shaped the adoption of Art NFTs and perception of their value. Third, he identifies the ideological and symbolic drivers of adoption. Some of these drivers for Art NFTs we have mentioned above, however he names more universal ones also that apply broadly across the Web3 ecosystem. These include the influence of cyber libertarianism, the mirage and allure of imagined futures, and the normalisation of practices and communication of ideologies that facilitate technology adoption.

Drawing on the concept of assemblages, Childs (2023) sees cryptomarkets as human and non-human forces in a dynamic system that is always in a process of becoming. In witness to the focus of his work, they are also always teetering on the brink of abrupt decline. We propose to extend this thinking by arguing that this decline for Art NFTs occurs when external forces of culture, social backlash and regulatory approaches snuff out the cruel optimism (Berlant, 2011) of a community. This is more than the disillusionment and disappointment that we see in technology hype cycles arising from the misfit of the sociotechnical imaginary with its materialisation (van Lente et al., 2013). We extend more firmly

[1] In the context of NFTs, whole of life marketplaces refer to platforms that support the entire lifecycle of a digital asset, from creation and minting to trading and retiring. These marketplaces typically allow users to create, buy, sell, and trade NFTs, as well as track ownership history and manage royalties. Examples include OpenSea, Rarible, and SuperRare.

[2] In-game economies are virtual economic systems within video games or virtual worlds where players can earn, trade, and spend virtual currencies or assets. These economies often involve the exchange of in-game items, currencies, or services, which may have real-world value. With the integration of blockchain technology and NFTs, some in-game economies have expanded to allow players to own, buy, and sell unique digital assets that can be used both within and outside the game environment.

upon these explanations of failure to say that the social licence to operate for this emerging technology has been revoked.

4. Hype and hustle

NFTs exploded in popularity in mid 2021 as pictures of apes sold for tens of millions of dollars, and an endless supply of headlines about million-dollar hacks of NFT projects and corporate cash grabs piled on top of each other (Clark, 2022). Here we see the bedfellows of hype and hustle in action. Understanding how the hope outlined in the previous section is transformed into and by hype and hustle, is an important part of understanding the social and cultural lives of NFTs as material objects, beyond the insider bubble and the influence of vested interests. The notion of hype refers to widely varied 'hype patterns' that explain hype cycles in which waves of media attention are combined with high expectations on technological possibilities and the associated attraction of attention, resources, coordination of activities, and the spurring of competition (van Lente et al., 2013). A common referent for this cycle is the Gartner hype cycle which describes the cycle in the evocative language of a technology trigger, a peak of inflated expectations, a trough of disillusionment, a slope of enlightenment and a plateau of productivity (O'Leary, 2008). van Lente et al. (2013) observe that hypes are usually followed by disappointment when high expectations are not met by the actual outcome of innovative activity. They observe however that whilst hype cycles in public discourses are often seen as exaggerated, deceptive, misleading, and presenting faulty predictions of the future, hype can also be seen as collectively pursued explorations of the future that have a performative capacity that affects activities in the present.

The seamless transition of Web3 hype cycles into the Web 2.0 and social media influencer landscape is unsurprising, given the digital, distributed, and peer-based nature of the ecosystem. NFTs gained prominence through celebrification, aligning with their platformisation in marketplaces like OpenSea, facilitating broader accessibility and social visibility. Influencers, particularly celebrities, play a crucial role in accelerating the adoption and legitimisation of emerging technologies, leveraging their social media presence. The hustle economy, as described by Cottom (2020: 19), involves influencers developing personal brands on social media platforms. Research indicates that social media influencers significantly impact the adoption of emerging technologies, enhancing consumers' intention to adopt endorsed applications through trust transfer (Hu et al., 2019). Finfluencers, providing financial advice on cryptocurrencies via social media, marked the initial convergence of Web3 hustling with broader digital cultures. Recent trends involve celebrities endorsing/boosting crypto products without properly disclosing that they were paid for their endorsements (Contreras, 2023). This section explores how celebrities and social media users employ hustle tactics to shape the social licence of Art NFTs, asserting that this mechanism determines the success or failure of the technology.

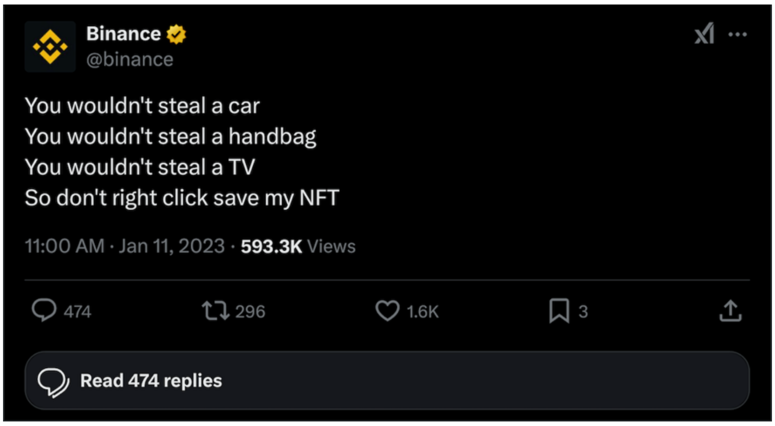
The concept of a social licence has circulated in academic literature since the 1970s (Dennis, 1975), and typically indicates the social permissibility of a particular behaviour, particularly where social norms might supersede legal ones. However, the term social license to operate emerges as an extension of this concept and is most frequently associated with mining and other resource-intensive industries (Gehman et al., 2017). Gehman et al. (2017) identify several models that scholars have used to theorise the social license to operate, but they argue what links them together is the concept of legitimacy. Legitimacy as a social concept can be traced back to Weber (1978), who links legitimacy to conformity to both formal and informal social norms. Scott (1995: 45) argues that "Legitimacy is not a commodity to be possessed or exchanged but a condition reflecting cultural alignment, normative support, or consonance with relevant rules or laws." Thus, the legitimacy or the social licence to operate that both NFTs and cryptocurrency rely on depends on a cultural alignment with their product offering, in this case ahead of relevant rules or laws. As identified earlier in this paper, NFTs and cryptocurrencies tend to initially operate in a legal grey zone where policy and regulation is yet to catch up with the novel actions and opportunism emerging technologies like this make possible.

The formalisation of the social license to operate through social scientific research confers the concept with a much more solid and structured form than is often the case. Discussions of the social licence to operate invoke formalised discussion between the community, key stakeholders, and the corporations, which reflect its origins in the resource industry (Owen and Kemp, 2013). While the social license to operate and legitimacy are two separate concepts, Gehman et al. (2017) note that they overlap considerably in their definition. We argue that the social licence to operate where it concerns new digital technologies, hinges on a level of permissibility that is predicated on the extent to which these technologies are culturally aligned with dominant social norms, and facilitated by what Aldrich and Fiol (1994: 648) identify as sociopolitical legitimacy, or "the process by which key stakeholders, the general public, key opinion leaders, or government officials accept a venture as appropriate and right, given existing norms and laws." It is the 'key opinion leaders' that we now turn our focus to, as they have been central in establishing NFT's social licence to operate.

Celebrities play a crucial role in conferring legitimacy to Non-Fungible Tokens (NFTs), leveraging their influence through endorsements and branded offerings. Notable figures, such as Paris Hilton and Tony Hawk, entered the NFT market during its peak, enhancing the socio-political legitimacy of these digital assets. Hilton is on the record publicly praising NFTs as an investment, stating that "I just started looking into what they were doing and was like, 'Wow these guys really know what they're doing, this is next level. I want to be involved, can I invest into this?'" (Youshaei, 2022). Beyond public statements of support, public displays of high-value NFT purchases, exemplified by Eminem and Snoop Dogg at the MTV Video Music Awards, further contributed to the perception of NFTs as a viable investment. Socialite turned crypto promoter Paris Hilton, was an investor in at least one NFT platform, and sold her own Planet Paris NFTs for more than $1 million (Wilser, 2021). Paris Hilton's dual role as a crypto investor and NFT seller, as observed by Mull (2022), underscores the speculative nature of NFTs, reliant on public investment for sustained growth. The traditional framework of celebrity endorsements falls short in comprehending the dynamics of NFT ownership and who profits. Mull (2022) contends that the influx of capital into cryptocurrency startups, often shrouded in secrecy, complicates the understanding of these celebrity-driven transactions. Gottsegen (2022) highlights the less visible influence of entities like MoonPay, a crypto custodian, in leveraging celebrity endorsements for marketing purposes. This practice, labelled as 'perverse deal-making', creates an illusion of an NFT gold rush, fostering FOMO and reinforcing individualistic, capitalistic neoliberal ideals amid societal uncertainties. It impresses upon individuals that they are responsible for their own financial wellbeing in what are increasingly difficult conditions and that they should "jump now" to shore up their future against uncertain conditions. It is not an accident that NFTs and cryptocurrencies boomed during the pandemic. But this social license to operate is predicated on NFTs' ability to hold ever-increasing value.

The social licence to operate is so central to NFTs because the financial backbone they rely on (cryptocurrency) is characterised as a trustless system in which the middleman (the bank or art broker/dealer/gallery) is removed. This removes a central credibility marker; the middleman and trust must therefore be placed in the technology. Arguably, blockchain infrastructure dispenses with the need to trust other people or, indeed, institutions (Dodd, 2018). Dodd (2018: 37) argues that rather paradoxically, the communities that have emerged around Web3 are sustained by the belief that these technologies (including Bitcoin) have "replaced social relation- the trust on which all forms of money depend-with machine code". However, following Dodd, we argue that NFTs thrive precisely because, due to the celebrity endorsement, and broader mainstreams of NFTs, people were prepared to trust in these new forms of technology, extending them a social licence to operate and social legitimacy in the absence of broader material benefit. This is essentially the influencer trust transfer process rather than trust in the technology. NFTs initial premise drew on utopian sociotechnical imaginaries that are also embedded in blockchain and of self-governance, financial freedom, and a monetary system uncoupled from the nation-state (Dodd, 2018). However, as highlighted above Web3 technologies, including NFTs, are very much reliant on existing social institutions, which both constrain and enable their capacity.

To a certain extent the 'hustle' of NFTs, their massification through celebrity culture, confused visibility with social legitimacy and a social licence to operate. While NFTs were hyper-visibly embraced by professional sports leagues, rappers and a wide variety of other celebrities, this phenomenon is not the same as legitimacy. While celebrity endorsements mainstreamed the concept and recognition of NFTs as a way of trading and storing value, it also ran contrary to other internet cultures and attitudes. For example, on Twitter NFTs became highly trollable. Briefly, Twitter provided special avatar frames to users who had an NFTs as their profile picture, distinguishing them from 'regular' profile pictures (Adams, 2024). While their purchase is recorded on the blockchain, the actual image of the NFT is still infinitely copyable, which other internet users could, and did do (Morse, 2021). Binance responded to this practice in a Jan 11 2023 tweet, in the style of the "you wouldn't steal a car" meme (Wikipedia, 2023), writing, "You wouldn't steal a car. You wouldn't steal a handbag. You wouldn't steal a TV. So don't right click save my NFT." (binance, 2023).



Binance
@binance

You wouldn't steal a car
You wouldn't steal a handbag
You wouldn't steal a TV
So don't right click save my NFT

11:00 AM · Jan 11, 2023 · 593.3K Views

Figure 1. Binance 'Right click save' tweet (Jan 11 2023)

While this tweet may not be completely serious, it illustrates the issues with generating broad-based social legitimacy for a digital ownership model that is immaterial and somewhat esoteric. The enrolment of celebrities to 'hustle' NFTs was one way in which the NFT industry attempted to close this gap. It is difficult to ascertain the precise networks through which this occurred, but the intermingling of celebrity and Web 3 industries are evident in NFT projects. For example, 'Stoner Cats', an NFT funded animation series helmed by Mila Kunis, has Ethereum co-creator Vitalik Buterin as the voice of one of the characters, Lord Catsington (Hayward, 2021). A lawsuit filed against Yuga Labs (most famous for the Bored Ape Yacht Clubs NFTs) suggests that Yuga Labs paid for celebrity endorsements and made them look 'organic' as a way of boosting prices (Whiddington, 2022). In using celebrities, money and visibility to establish a perception of socio-political legitimacy, the initial hope of NFTs, that they would help support more equitable models of income for artists was drained.

While artists may have benefited from the boom of NFTs in some ways, the copy-paste affordance of digital content also meant that some had their work copied and sold as NFTs without their permission (ArtsLaw, 2024), placing the income stream in the hands of those who were not the creators of the content. The hope of Art NFTs is further undermined by the shifting of their purchase and sale off the blockchain. As discussed above, the blockchain upon which both cryptocurrencies and NFT rest, is meant to be a 'trust free' technology, erasing their need for a meditating party through which these transactions are managed. However, the blockchain itself is slow, often expensive to interact with, and riddled with adhoc coding that ultimately makes it easy to steal NFTs, which people did. To extend the cultural reach of NFTs then, platforms and corporations developed intermediary interfaces to facilitate buying, selling and collecting, for example collectible NFTs sold by the NBA were exchanged through their platform, Starbucks also supplied the platform infrastructure to support its (failed) NFT venture. What is an NFT without the utopian ideal of trust-free technology underpinning frictionless exchange? It's a hustle.

5. Death and taxes

In this section, we consider the death throes of Art NFTs in the media, and from environmental, reputational, and legal issues and regulatory intervention, resulting in their loss of a social licence to operate. Media coverage and industry reports over 2023 collectively portray a troubled landscape for the NFT market, with numerous reports indicating a substantial devaluation of NFTs. For example, NFT sales were reported to have a sharp decline, peaking at USD 12.6 billion in January 2022 and reaching barely over USD 1 billion by June 2022 (Milmo, 2022). By October 2022, sales had fallen more than 90% from the previous year's measurements in nearly every category, including volume and price (Parisi, 2022). Despite these hurdles, the market showed signs of a post-crash recovery however the DappRadar data from October 2023 indicates a continued decline in NFT market indicators and also a rise in exploits and hacks within the decentralised application sector (Gherghelas, 2023). We observe that it is no accident that most of the legal and regulatory changes in response to NFTs and cryptocurrency have occurred in what is called the 'crypto winter' the vernacular for the massive loss in market cap after the crypto and Web3 bubbles burst. The significant drop in the value of NFTs has had financial repercussions for collectors, artists, and auction houses, as highlighted in various cases and lawsuits.

The complex nature of NFT investments, coupled with legal uncertainties, poses challenges for the long-term sustainability of the market. In an early observation on these issues that have dogged crypto projects since their inception, Smith (2019) argued that there was a clear need for a critical re-evaluation of NFT investments and a deeper understanding of the market dynamics to navigate its uncertainties successfully. However regulatory approaches towards NFTs are more likely to follow the crypto regulatory playbook that sees the implementation of chokepoints or a bottleneck strategy of governance (Smith, 2019) through mechanisms such as taxation and regulatory classification - property or currency, asset or security, legal or illegal, innovative financial technology or tool for criminal activity. An example of a recent but extended classification roadblock can be found in the actions of the US Securities and Exchange Commission who is reportedly "a bogeyman for the crypto industry" and rejected many applications for Bitcoin to be offered as an ETF (Exchange-Traded Funds) since 2013 for example (Dugan, 2024). The main concerns here for regulators being about consumer protections against market manipulation and investors losing money. This has meant, Dugan (2024: np) argues, that to date there has not been an "easy, cheap or low-risk way for regular folks with a 401(k) or a brokerage account to buy into the digital currency", stymieing its mainstreaming within the investment sector. While Bitcoin and Ether have now gained ETF status (Zaslowsky, 2024), this bottlenecking strategy of governance acts in lieu of outright banning/throttling remains - however Smith (2019) observes that it is difficult to implement for a decentralised and easily replicable technology. The intricate interplay between legal and financial mechanisms create potential avenues through which regulatory bodies could exert control despite the decentralised nature of the technology.

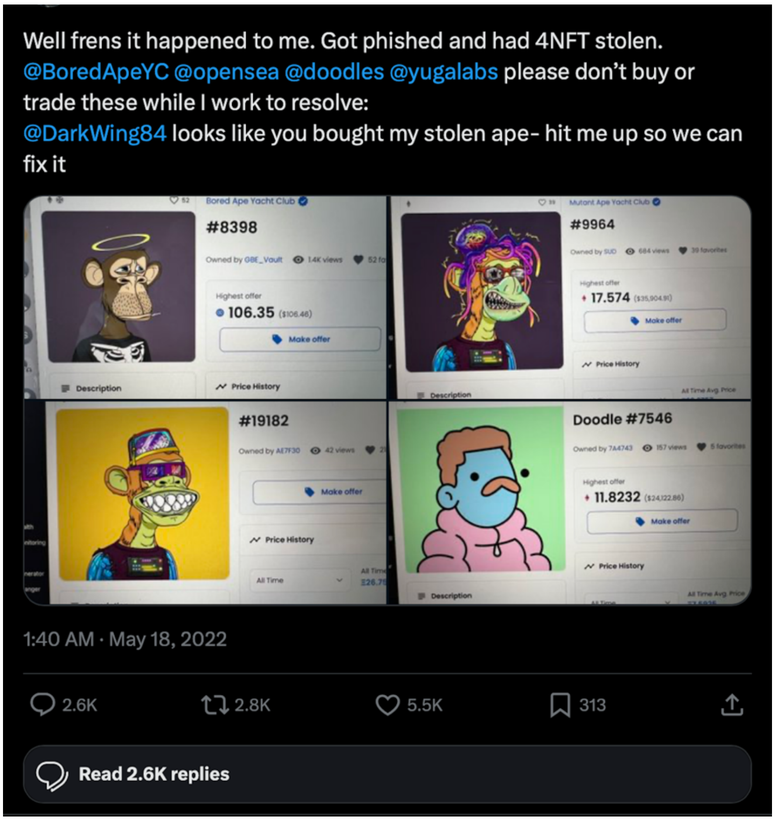
Given their decentralised and digital-only presence, these technologies are de-materialised in popular imagination. Existing on the blockchain and within the Web3 ecosystem, they often seem untethered from the materialities of technologies, interfaces, and production. This obscuring of the material conditions of NFT production makes it easier for consumers to buy into the utopian dream. However, like all internet-based technologies (Velkova, 2019), they do have a very material infrastructure through their reliance on cloud computing and more uniquely, the computing power they require to be minted (produced onto the blockchain). The White House reported that in 2022 that cryptocurrency assets exceeded national electricity usage, accounting for 0.4% to 0.9% of the world's yearly electricity consumption (OSTP, 2022). However, the exact means through which the calculation of energy consumption of blockchain-based technologies appears to be inconsistent (BBC, 2017), leading to claims for example that the Bitcoin

https://doi.org/10.33621/jdsr.v7i154881

cryptocurrency consumes as much power as the nation-state of Ireland (Hern, 2017) and to the refutation of this claim (Bevand, 2017). Of the principle behind this issue, Giungato et al. (2017) elaborate that the system upon which both fungible and non-fungible tokens are generated has been built in a way almost like the mining of a natural resource: costs and efforts rise as the system reaches the ultimate resource limit. The verification or consensus mechanism through which the intensification of resource consumption is achieved on the blockchain is referred to as Proof of Work. Within the broader discourse on the environmental consequences of cryptocurrencies and blockchain technologies, an ongoing critique scrutinises the environmental implications of Non-Fungible Tokens (NFTs), particularly within the realm of 'cryptoart'.

This scrutiny of Art NFTs, often characterised in media narratives (Calma, 2021) and by activist artist collectives and scholars (Calvo, 2023; Truby et al., 2022) as environmentally unethical, has generated claims of threats to global temperature and increased death rates through energy consumption and emissions that, while resting on inconsistent measurement approaches, bear reputational repercussions regardless. In review of the wide-ranging commentary, it appears that non-fungible tokens (NFTs) may be harmful to the environment depending on how they are produced aka minted (Garnett et al., 2022). The blockchain underpinning many NFTs, Ethereum, used Proof of Work up until mid-2022 which was an environmentally costly process and was the focus of much of this critique. NFT platform providers such as Tezos positioned themselves early within this ecosystem as an environmentally sustainable blockchain that contrasted to Ethereum based upon its use of the Proof of Stake consensus mechanism (Evans, 2023; Segre, 2023). Ethereum shifted to proof of stake in September 2022, a move touted by Ethereum's founder, Vitalik Buterin to reduce its global electricity usage by 0.2% and cut crypto carbon emissions by 99.992% (Vitalik.eth, 2022). Truby et al. (2022) argue that social pressure from the art market prompted the switch away from resource hungry proof-of-work blockchains to more sustainable consensus protocols, however commensurate social pressure, and the community itself have been working on this issue for much longer. Despite this change, it remains unclear whether Ethereum's PoS really is a sustainable alternative to PoW (Ho, 2023) and it appears that a deliberately high energy-intensive Proof-of-Work blockchain remains the most popular choice for blockchain consensus protocols (Truby et al., 2022).

Beyond their environmental concerns, NFTs remain problematic from a legal stance. The predominant legal challenges associated with Non-Fungible Tokens (NFTs), as posited by Jia and Yao (2024), primarily revolve around issues concerning the attribution and utilisation of Intellectual Property (IP) rights pertaining to the underlying content. Furthermore, they observe that legal disputes often arise in the form of non-contractual matters, such as instances of theft. Illustrating the convergence of intellectual property (IP) and theft concerns, the case of a BAYC NFT owned by actor and creator Seth Green serves as an example. In 2021, Green obtained Bored Ape Yacht Club NFT #8398, and dedicated substantial efforts to creating the series "White Horse Tavern" based on this NFT as the main character, only to have it and 3 other Bored Ape NFTs stolen through a phishing scam (Newar, 2022) and then on sold as a part of a larger multi-million-dollar scam operation (Emerson, 2022a).



Well frens it happened to me. Got phished and had 4NFT stolen. @BoredApeYC @opensea @doodles @yugalabs please don't buy or trade these while I work to resolve:
@DarkWing84 looks like you bought my stolen ape- hit me up so we can fix it

1:40 AM · May 18, 2022

Figure 2. Seth Green's Tweet (18 May 2022) indicating his Bored Ape NFTs were stolen through a Phishing scam.

The theft raised questions on whether Green was still allowed to use the Bored Ape for the show or if he lost the BAYC intellectual property rights that the NFT came with once it was stolen and on sold (Rizzo, 2022). BAYC's license does not stipulate instances of theft but states simply that when purchased, the NFT smart contract terms mean the owner holds the underlying Bored Ape, the Art, completely (Newar, 2022). Interpretations of these terms varied in the resulting commentary surrounding the theft, with some believing this meant that even if the NFT is bought from a thief, the usage rights transfer to the new owner. The absence of established legal precedent in this matter necessitated a hiatus in the show's development (Zilko, 2022), thereby exemplifying the real-world impacts of the unresolved complex legal challenges inherent in litigation pertaining to stolen NFTs (Newar, 2022). Green indicated publicly that he would go to court to get back the BAYC NFTs if it was not returned by buyer (Green, 2022). The stolen NFTs were eventually returned to Green for a reported $100,000 premium (Emerson, 2022b). More recent instances of NFT related cyberattacks, such as the NFT Trader attack through smart contract vulnerabilities (CryptoNews, 2023), highlight the ongoing risks associated with speculative investments in digital assets amid evolving cyber threats. Further complicating such matters, Jia and Yao (2024) point out that legal cases involving NFTs frequently manifest an international dimension, given the decentralised nature of the technology underpinning their development. This decentralisation is reflected in their distribution across servers located in numerous countries, coupled with the global user connectivity facilitated by trading platforms, thus contributing to the transnational nature of litigations in this domain.

A renewed focus on regulation of the Web3 space has been further spurred by the entanglement of key crypto and NFT figures with the criminal justice system. IMF backed commentary provided in July 2023 on the emerging challenges governments face of taxing crypto assets notes that:

https://doi.org/10.33621/jdsr.v7i154881

> The collapse of FTX last year and recent US Securities and Exchange Commission lawsuits against Binance and Coinbase have fed anxiety among users while the appeal to criminal activities has been reflected in high-profile seizures of billions of dollars. These developments have triggered increasing scrutiny from policymakers and widespread calls for regulation (Baer et al., 2023: np).

In November 2023, one of the key figures associated with the crypto boom and bust, Sam Bankman-Fried was convicted of fraud and conspiracy in the Manhattan Federal Court (Cohen and Godoy, 2023). Likewise, the CEO of Binance, a crypto trading platform, also pled guilty on November 21. In the NFT space, the developer of the Mutant Ape Planet NFT project was charged in January 2023 with a 2.9 million dollar 'rug pull' and subsequently pled guilty (USAO, 2023). A rug pull is a scam in the cryptocurrency or NFT space where developers encourage investors to buy into the project, then abandon it and abscond with the invested funds, leaving the digital tokens or NFTs essentially worthless. Michel allegedly promised investors exclusive rewards, giveaways, and access to a marketplace for NFTs, but never delivered on these promises. Instead of continuing to develop the project, Michel allegedly transferred the funds to his personal wallets. The NFTs didn't become entirely worthless immediately, but their value significantly decreased due to the lack of promised development and benefits. Regardless, it is difficult to argue for the success of 'trust-free' technologies when they are litigated in court. Web3 is meant to be self-governing and self-regulating; every failure puts to light that this is not a functioning proposition. It is the community that (in part) provides the legitimacy for NFTs, and the community that can withdraw its agreement to support their social license to operate. In death we can observe the intervention of legal and regulatory frames which may have previously struggled to catch up to surging new technologies. Part of the imposition of legal and regulatory frames is due to the failure of the social licence to operate.

Like most emerging technologies, Art NFTs rely on network adoption, a conducive or slow-to-catch-up regulatory environment, and an underlying community that tinkers with the possibilities they offer, expresses creativity and entrepreneurial endeavour through them, and imbues them with perceived value (seeing as how tokens on a blockchain have no material or otherwise inherent value). Lotti (2016: 105) argues that normative power of the blockchain alone is not enough to emancipate art from contemporary financial logic. Further arguing that tokenisation can reproduce and amplify existing financial logics in the digital sphere by offering more precise ways to monetise digital interactions and take advantage of the speculative nature of markets (Lotti, 2016: 288). We also observe a cultural clash between hobbyist logic and capitalist cultures within NFT communities, showcasing the challenges faced by technologies seeking long-term persistence. As Calvo (2023: np) points out in what is clearly a provocative position piece on Art NFTs there are three principal positions that those in the art world take towards cryptoart: "those who believe it is a new bubble; those who think it is a revolution; and those who think the idea has failed". We observe that hobbyist cultures persist after failure in practices of creation, while speculative capitalist tech cultures move on to the next thing. Already it is clear that the pillars of NFT development and visibility, venture capital money, social media platforms and talent, have shifted from crypto-projects to (Generative) AI (cf. Coll-Beswick, 2023).

6. A eulogy for Art NFTs

This article has proposed a 'lifecycle' model through which we might understand the hope, hype and hustle, death and failure cycle of Art NFTs. In this concluding section, we reflect on the 'death' of Art NFTs as a mainstreamed social phenomenon and consider the broader applications for the technological emergence lifecycle model with this downward arc built in.

We acknowledge that Art NFTs still have a place in the broader art ecosystem, but we contend that these remain 'edge' cases, and the sociotechnical imaginary filled with hope, hype and hustle that fuelled the boom of Art NFTs, now lacks social legitimacy and social licence. Sociologically finessed ideas of legitimacy and a social licence to operate have rarely been extended to the sociotechnical sphere in focus in this article. However, as we have demonstrated in our analysis of Art NFTs, they play an important role in the lifecycle of emerging technologies, and often function ahead of institutional regulatory responses.

Art NFTs were not an innovation without demand. They are responding to legitimate issues that artists face in response to the difficulty of making a living through art. Art NFTs open a potentially transformational relationship to art creation, sale, and resale. However, as most recently explore in the documentary *The Stormtrooper Scandal* (Mangan, 2024), these exchanges can quickly become ethically and practically murky. In this instance (and many similar as previously enumerated), the potential financial rewards of NFTs quickly outstripped any broader ideological project or utility to artists. Ultimately, the Stormtrooper NFT project culminated in a quasi-rug-pull. The NFTs had been minted without the artist's permission, and those involved in the crypto side of the project quickly disappeared into the digital wilds after receiving their cut. With the NFTs eventually removed from sale by OpenSea, the investors, and the creator of the project, Ben Moore, have been left holding the (empty) bag.

To unpack the implications of our findings, it is essential to consider the broader social and economic context in which the Art NFT phenomenon emerged and declined. NFTs boomed in the precarious economic conditions of the COVID-19 pandemic and seemed to promise a way out of the economic instability of late-capitalism, compounded by the uneven impacts of the pandemic. Art NFTs, and arguably the broader Web 3 system (including blockchain and cryptocurrency) seemed to offer a new democratised model of investment, and (potential) financial freedom in the face of growing precarity and inequality.

The financial incentives of Art NFTs were also supported by their perceived cultural and social legitimacy. The lifecycle described through this study owes a lot of its intensity to the involvement of celebrity, which brought Art NFTs broad, mainstream legitimacy. Further, cultural gatekeepers like Christie's leant their institutional reputation to the Art NFT project, where at its peak they auctioned Beeple's *Everydays – The First 5000 Days* for auction, ultimately selling for $69 million. While Christie's NFT project continues, Sotheby's a competing auction house, also involved in the art NFT space is being sued for allegedly deceptively representing the level of mainstream interest in Bored Ape Yacht Club NFTs (Chow, 2023). These issues, among a slew of others too numerous to discuss fully in this space, illustrate the long tail of the 'death' of emerging technologies, and raise attendant questions about who ultimately bears responsibility for these risks.

Through our conceptual framing of the Art NFT hope, hype, hustle, and failure lifecycle, we propose a real-world derived model for analysing technological trajectories and potential failures. This model offers valuable insights beyond the realm of Art NFTs, proving applicable to various emerging technologies. The cycle of hope, hustle, hype, death, and taxes that we've observed in the Art NFT space are a recurring pattern in the broader technological landscape.

By applying this lens to other innovations, we can better illuminate and acknowledge the generative tensions between technological potential, societal expectations, and practical realities. Consider, for instance, the 2021/2022 upwelling of investment and hype around the metaverse (Chow, 2022; Grayscale, 2021), and Meta's subsequent shift away from this hype (Sevilla, 2022). Similarly, the current landscape of generative AI (GenAI) tools demonstrates a surge of hype and hustle reminiscent of the early stages of the Art NFT boom. The sociotechnical imaginary surrounding GenAI is rife with promises of revolutionary change across industries. However, as our model suggests, this prevalent hype tends to obscure a clear view of the technology's actual capabilities, limitations, and societal implications.

Our lifecycle model offers a real-world derived tool for analysing these emerging technologies. By identifying stages, anticipating challenges, guiding development, informing regulation, and encouraging critical analysis, this model can foster more informed, ethical, and socially responsible approaches to technological innovation. Understanding where in the lifecycle a technology sits can potentially help us make more socially responsible and ethical choices about its regulation and development.

References

(2022) *Frosties presale*. Available at: https://nftcalendar.io/event/frosties/ (accessed 12 July 2024).

Abutu JE (2023) *The rise of NFT Art in 2023: A deep dive.* Available at: https://johnedwinabutu.medium.com/the-rise-of-nft-art-in-2023-a-deep-dive-72372e0270d8. 13 September 2023. (accessed 15 January 2025).

Adams J (2024) Twitter NFT profiles quietly removed as hype fades into memory. *CNN*. 10 January 2024. Available at: https://www.ccn.com/news/twitter-nft-profiles-removed-hype-fades/ (accessed 15 January 2024).

Aldrich HE and Fiol CM (1994) Fools Rush in? The Institutional Context of Industry Creation. *The Academy of Management Review* 19(4): 645-670. https://doi.org/10.2307/258740.

Ang I (2021) Beyond the crisis: transitioning to a better world? *Cultural studies (London, England)* 35(2-3): 598-615. https://doi.org/10.1080/09502386.2021.1898013

ArtsLaw (2024) Non-fungible Token. *ArtsLaw*. Available at: https://www.artslaw.com.au/information-sheet/non-fungible-token-nft/ (accessed 15 January 2024).

Baer K, de Mooij R, Hebous S, et al. (2023) Crypto Poses Significant Tax Problems—and They Could Get Worse. In: IMF Blog. 5 July 2023.Available at: https://www.imf.org/en/Blogs/Articles/2023/07/05/crypto-poses-significant-tax-problems-and-they-could-get-worse (accessed 15 January 2025).

Barbrook R and Cameron A (1996) The Californian ideology. *Science as Culture* 6(1): 44-72. https://doi.org/10.1080/09505439609526455.

BBC (2017) Bitcoin: Does it really use more electricity than Ireland? *BBC*, 12 December 2017. https://www.bbc.com/news/technology-42265728 (accessed 15 January 2025).

Berlant L (2011) *Cruel Optimism.* Durham: Durham : Duke University Press.

Bevand M (2017) Serious faults in Digiconomist's Bitcoin Energy Consumption Index. In: mrb's blog. 1 February 2017. Available at: http://blog.zorinaq.com/serious-faults-in-beci/ (accessed 15 January 2025).

Beyer S (2023) Metamodern Spell Casting : The Blockchain as a Conceptual Medium for Contemporary Visual Artists. *M/C Journal* 26(5). https://doi.org/10.5204/mcj.2999.

Bickford A (2018) From Idiophylaxis to Inner Armor: Imagining the Self-Armoring Soldier in the United States Military from the 1960s to Today. *Comparative studies in society and history.* 60(4): 810-838. https://doi.org/10.1017/S0010417518000300.

binance (2023) You Wouldn't steal a car. 11:00AM ed.: X. 11 January 2023. https://x.com/binance/status/1612962619656118273.

BlockTides (2023) Featured NFT | Fidenza #313: A Historic Leap in Art Blocks' Ethereum Collection. *Medium*. 9 August 2023. Available at: https://blocktides.medium.com/featured-nft-fidenza-313-a-historic-leap-in-art-blocks-ethereum-collection-4a5d53863421 (accessed 30 August 2024).

Calma J (2021) The climate controversy swirling around NFTs. *The Verge*. 16 March 2021.Available at: https://www.theverge.com/2021/3/15/22328203/nft-cryptoart-ethereum-blockchain-climate-change (accessed 15 January).

Calvo P (2023) Cryptoart: Ethical Challenges of the NFT Revolution. *arXiv.org*. https://doi.org/10.48550/arxiv.2307.03194.

Carroll J, Howard S, Peck J, et al. (2003) From Adoption to Use: the process of appropriating a mobile phone. *AJIS. Australasian journal of information systems* 10(2). https://doi.org/10.3127/ajis.v10i2.151.

Carroll J, Howard S, Vetere F, et al. (2002) Just what do the youth of today want? Technology appropriation by young people. *35th Annual Hawaii International Conference on System Sciences (HICSS-35 2002).* Hawaii, 1777-1785. https://doi.org/10.1109/HICSS.2002.994089.

Chalmers D, Fisch C, Matthews R, et al. (2022) Beyond the bubble: Will NFTs and digital proof of ownership empower creative industry entrepreneurs? *Journal of Business Venturing Insights* 17: e00309. https://doi.org/10.1016/j.jbvi.2022.e00309.

Childs A (2023) How cryptomarket communities navigate marketplace structures, risk perceptions and ideologies amid evolving cryptocurrency practices. *Criminology & Criminal Justice* 0(0): https://doi.org/10.1177/17488958231213012.

Choi C (3023) 'Bored Apes' investors sue Sotheby's, Paris Hilton and others as NFT prices collapse. *CNN*. 17 August 2023. Available at: https://edition.cnn.com/style/article/bored-apes-sothebys-lawsuit/index.html (accessed 30 August 2024).

Chow A (2022) A Year Ago, Facebook Pivoted to the Metaverse. Was It Worth It? *Time*. 27 October 2022. Available at: https://time.com/6225617/facebook-metaverse-anniversary-vr/ (accessed 30 August 2024).

Chow V (2023) A Group of Collectors Is Suing Sotheby's Over Its 'Misleading' Marketing of Bored Ape Yacht Club NFTs. artnet.15 August 2023. Available at: https://news.artnet.com/art-world/collectors-sue-sothebys-bored-ape-yacht-club-nfts-2349974 (accessed 30 August 2024).

Clark M (2022) NFTs explained. *The Verge*. Updated 6 Jun 2022. Available at: https://www.theverge.com/22310188/nft-explainer-what-is-blockchain-crypto-art-faq (accessed 15 January 2024).

Cohen L and Godoy J (2023) Sam Bankman-Fried convicted of multi-billion dollar FTX fraud. *Reuters*. 3 November 2023. Available at: https://www.reuters.com/legal/ftx-founder-sam-bankman-fried-thought-rules-did-not-apply-him-prosecutor-says-2023-11-02/ (accessed 15 January).

Coll-Beswick C (2023) AI Is Killing Crypto Venture Capital Interest. *Coindesk*. Available at: https://www.coindesk.com/opinion/2023/09/11/ai-is-killing-crypto-venture-capital-interest (accessed 15 January 2025).

Contreras B (2023) Lindsay Lohan, Jake Paul, Lil Yachty, other celebs hit with SEC charges for boosting crypto. *Los Angeles Times*, 22 March 2023.

Cottom TM (2020) The Hustle Economy. *Dissent.* 67(4): 19-25. https://doi.org/10.1353/dss.2020.0094.

CryptoNews (2023) Million-Dollar NFT Heist: Apes to Art Blocks Stolen. *Crypto News*. 18 December 2023. Available at: https://www.cryptonews.net/news/nft/28253717/ (accessed 15 January).

Dennis PA (1975) The Role of the Drunk in a Oaxacan Village. *American Anthropologist* 77(4): 856-863. https://doi.org/10.1525/aa.1975.77.4.02a00080.

Dodd N (2018) The Social Life of Bitcoin. *Theory, Culture & Society* 35(3): 35-56. https://doi.org/10.1177/0263276417746464/.

Dugan KT (2024) Why Wall Street Might Be Falling in Love With Bitcoin. *New York Magazine*. 2 January 2024.Available at: https://nymag.com/intelligencer/2024/01/the-long-awaited-bitcoin-etf-is-probably-almost-here.html (accessed 15 January).

Elliptic (2022) *NF*Ts and Financial Crime. *Elliptic*. Available at: https://www.elliptic.co/resources/nfts-financial-crime (accessed 12 July 2024).

Emerson S (2022a) Seth Green's Stolen Bored Ape Is Back Home. *BuzzFeed News*. 10 June 2022. Available at: https://www.buzzfeednews.com/article/sarahemerson/seth-green-bored-ape-nft-returned (accessed 15 January).

Emerson S (2022b) Someone Stole Seth Green's Bored Ape, Which Was Supposed To Star In His New Show. *BuzzFeed News*. 24 May 2022. Available at: https://www.buzzfeednews.com/article/sarahemerson/seth-green-bored-ape-stolen-tv-show. (accessed 15 January).

Evans L (2023) Why Are NFTs Bad for the Environment? A Deep Dive into Tezos vs. Ethereum. *XTZ news*. 7 September 2023. Available at: https://xtz.news/en/opinion/why-are-nfts-bad-for-the-environment/ (accessed 15 January).

Garnett AG, Brown JR and Rohrs Scmitt K (2022, 2024) NFTs and the Environment: What You Need to Know. *Investopedia*. Updated 11 April 2024. Available at: https://www.investopedia.com/nfts-and-the-environment-5220221 (accessed 15 January 2025).

Gehman J, Lefsrud LM and Fast S (2017) Social license to operate: Legitimacy by another name? *Canadian Public Administration* 60(2): 293-317. https://doi.org/10.1111/capa.12218.

Gherghelas S (2023) State of the Dapp Industry Q3 2023. *DappRadar*. 5 October 2023.In: Available at: https://dappradar.com/blog/state-of-the-dapp-industry-q3-2023 (accessed 2024).

Gillis A (2022) *Hockey stick growth*. TechTarget. Last updated October 2022. Available at: https://www.techtarget.com/searchcustomerexperience/definition/hockey-stick-growth (accessed 15 January 2024).

Girder. D. (2021) The Metaverse: Web 3.0 Virtual Cloud Economies. Grayscale. Last update 24 November 2021. Available at: https://www.grayscale.com/research/reports/the-metaverse (accessed 15 January 2025).

Giungato P, Rana R, Tarabella A, et al. (2017) Current Trends in Sustainability of Bitcoins and Related Blockchain Technology. *Sustainability.* 9(12): 2214. https://doi.org/10.3390/su9122214.

González RJ (2022) *War virtually : the quest to automate conflict, militarize data, and predict the future.* Oakland, California : University of California Press.

Gottsegen W (2022) *NFTs, celebrities and perverse deal-making*. CoinDesk. 31 January 2022, Updated 14 June 2024 Available at: https://www.coindesk.com/opinion/2022/01/31/nfts-celebrities-and-perverse-deal-making (accessed 15 January 2025).

Green S (2022) Well frens it happened to me. In: @SethGreen 1:40AM 18 May 2022. *X*. Available at: https://x.com/SethGreen/status/1529187693984329728 (accessed 15 January 2025)

Green S (2022) Not true since the art was stolen. A buyer who purchased stolen art with real money and refuses to return it is not legally entitled to exploitation usage of the underlying IP. It'll go to court, but I'd prefer to meet @DarkWing84 before that. Seems we'd have lots in common. In: @SethGreen 5:48AM 25 May 2022. *X*. Available at: https://x.com/SethGreen/status/1526588358859759617 (accessed 15 January 2025).

Grimmelmann J, Ji Y and Kell T (2022a) Copyright vulnerabilities in NFTs. In: The Initiative for CryptoCurrencies and Contracts (IC3). *Medium*. 22 March 2022. Available at: https://medium.com/initc3org/copyright-vulnerabilities-in-nfts-317e02d8ae26 (accessed 15 January 2025).

Grimmelmann J, Ji Y and Kell T (2022b) The tangled truth about NFTs and copyright. *The Verge*, 8 June 2021. Available at: https://www.theverge.com/23139793/nft-crypto-copyright-ownership-primer-cornell-ic3 (accessed 15 January 2025).

Harris E (2023) Mint, sell, repeat: Non-fungible tokens and resale royalties for Indigenous artists. *Alternative law journal.* 48(1): 11-16. https://doi.org/10.1177/1037969X221141096

Hayward A (2021) Mila Kunis' 'Stoner Cats' Cartoon Is Making Millions Selling NFTs. *Vice*. 29 July 2021. Available at: https://www.vice.com/en/article/mila-kunis-stoner-cats-cartoon-is-making-millions-selling-nfts/ (accessed 30 August 2024).

Hedera (2023) *W*hat is an NFT Smart Contract? *Hedera.* Available at: https://hedera.com/learning/smart-contracts/nft-smart-contract (accessed 15 January 2024).

Hendry N, Hanckel B and Zhong A (2021) *Navigating uncertainty: Australian young adult investors and digital finance cultures.* RMIT University. https://doi.org/10.25916/zbje-qn11.

Henry P (2018) How to explain your startup's 'Hockey Stick' revenue growth without appearing naive. In: *LinkedIn*. 1 March 2018. Available at: https://www.linkedin.com/pulse/how-explain-your-startups-hockey-stick-revenue-growth-patrick-henry/ (accessed 2024).

Hern A (2017) Bitcoin mining consumes more electricity a year than Ireland. *The Guardian*, 27 November 2017. Available at: https://www.theguardian.com/technology/2017/nov/27/bitcoin-mining-consumes-electricity-ireland (accessed 15 January 2025).

Ho C (2023) One Year After The Merge: Sustainability Of Ethereum's Proof-Of-Stake Is Uncertain. *Forbes*. 11 October 2023. Available at: https://www.forbes.com/sites/digital-assets/2023/10/11/one-year-after-the-merge-sustainability-of-ethereums-proof-of-stake-is-uncertain (accessed 15 January).

Hobbs T (2021) Fidenza. Blog: *TylerHobbs*. 16 November 2021. Available at: https://www.tylerxhobbs.com/words/fidenza (accessed 30 August 2024).

Houlbrook-Walk M (2022) Indigenous artists from East Arnhem Land venture into billion-dollar NFT digital marketplace. *ABC News*, 21 March 3033. Available at: https://www.abc.net.au/news/2022-03-21/yolngu-artists-yirrkala-nfts/100916188 (accessed 15 January 2025).

Howcroft E and Carvalho R (2021) Insight: How a 10-second video clip sold for $6.6 Million. *Reuters*, 1 March 2021. Available at: https://www.reuters.com/business/media-telecom/how-10-second-video-clip-sold-66-million-2021-03-01/ (accessed 15 January 2025).

Hu H, Zhang D and Wang C (2019) Impact of social media influencers' endorsement on application adoption: A trust transfer perspective. *Social Behavior and Personality: an international journal* 47(11): 1-12. https://doi.org/10.2224/sbp.8518.

Jasanoff S and Kim S-H (2015) *Dreamscapes of modernity : sociotechnical imaginaries and the fabrication of power.* Chicago London The University of Chicago Press.

Jia W and Yao B (2024) NFTs applied to the art sector: Legal issues and recent jurisprudence. *Convergence* 30(2): 807-822. https://doi.org/10.1177/13548565231199966.

Kaur, G. (2023) A beginner's guide on the legal risks and issues around NFTs. *Coin Telegraph*. 15 August 2023. Available at: https://cointelegraph.com/learn/a-beginners-guide-on-the-legal-risks-and-issues-around-nfts (accessed 15 January 2024).

Kelso A (2023) Digital artist Ben Fowler tells how NFTs create a new global marketplace for Australia's regional creatives. *ABC News*. 23 September 2023. Available at: https://www.abc.net.au/news/2023-09-23/ben-fowler-making-a-name-for-himself-digital-art-nfts/102848890 (accessed 30 August 2024).

Lotti L (2016) Contemporary art, capitalization and the blockchain: On the autonomy and automation of art's value. *Finance and Society.* 2(2): 96-110. https://doi.org/10.2218/finsoc.v2i2.1724

Mackenzie S and Bērziņa D (2022) NFTs: Digital things and their criminal lives. *Crime, media, culture.* 18(4): 527-542. https://doi.org/10.1177/17416590211039797.

Mangan L (2024) The Stormtrooper Scandal review – inside the Star Wars art sale that wrecked lives. *The Guardian*. 21 June 2024. Available at: https://www.theguardian.com/tv-and-radio/article/2024/jun/20/the-stormtrooper-scandal-review-inside-the-star-wars-art-sale-that-wrecked-lives (accessed 30 August 2024).

McMillan M, Lindhout R and Morgan K (2020) *Smart(er) contracts in 2020*. Mondaq. 9 August 2020. Available at: https://www.mondaq.com/australia/new-technology/974460/smarter-contracts-in-2020 (accessed 15 January 2025).

Merton RK (1936) The Unanticipated Consequences of Purposive Social Action. *American Sociological Review* 1(6): 894-904. https://doi.org/10.2307/2084615.

Milmo D (2022) NFT sales hit 12-month low after cryptocurrency crash. The Guardian. *The Guardian*, 2 July 2022. Available at: https://www.theguardian.com/technology/2022/jul/02/nft-sales-hit-12-month-low-after-cryptocurrency-crash (accessed 25 January 2025).

Morse J (2021) NFT owners insist they're totally not owned by 'right-click savers'. *Mashable*. 18 August 2021. Available at: https://mashable.com/article/non-fungible-tokens-nfts-right-click-save (accessed 15 January 2024).

Mull A (2022) Celelbrities and NFTs are a match made in hell. *The Atlantic*, 4 February 2022. Available at: https://www.theatlantic.com/technology/archive/2022/02/nft-jimmy-fallon-paris-hilton-millionaire/621486/ (accessed 15 January 2025).

Nadini M, Alessandretti L, Di Giacinto F, et al. (2021) Mapping the NFT revolution: market trends, trade networks, and visual features. *Scientific reports* 11(1): 20902. https://doi.org/10.1038/s41598-021-00053-8.

Neves BB, Waycott J and Maddox A (2023) When Technologies are Not Enough: The Challenges of Digital Interventions to Address Loneliness in Later Life. *Sociological Research Online* 28(1): 150-170. https://doi.org/10.1177/13607804211029298.

Newar B (2022) 'Code is not law:' Seth Green thief stole Bored Apes, not the rights, say experts. *Cointelegraph*. 25 May 2022. Available at: https://cointelegraph.com/news/code-is-not-law-seth-green-thief-stole-bored-apes-not-the-rights-say-experts (accessed 15 January 2025).

NFTevening (2023) Fidenza Incomplete Control Sells $7 Million worth of NFTs Before the Reveal. *NFTevening*. Updated 25 November 2024. Available at: https://nftevening.com/fidenza-incomplete-control-sells-7-million-worth-of-nfts-before-the-reveal/ (accessed 30 August 2024).

O'Leary DE (2008) Gartner's hype cycle and information system research issues. *International Journal of Accounting Information Systems* 9(4): 240-252. https://doi.org/10.1016/j.accinf.2008.09.001.

OSTP (2022) Climate and energy implications of crypto-assets in the United States. *White House Office of Science and Technology Policy*. Washington, D.C. September 8, 2022. Available at: https://www.whitehouse.gov/wp-content/uploads/2022/09/09-2022-Crypto-Assets-and-Climate-Report.pdf

Owen JR and Kemp D (2013) Social licence and mining: A critical perspective. *Resources policy* 38(1): 29-35. https://doi.org/10.1016/j.resourpol.2012.06.016.

Parisi D (2022) 2022 Was the Year of the NFT Reality Check. *Glossy.* 27 December 2022. Available at: https://www.glossy.co/fashion/2022-was-the-year-of-the-nft-reality-check/ (accessed 15 January 2024).

Pereira AP (2023) *NFT Trader hacked, millions of dollars in NFT stolen*. Cointelegraph. 16 December 2023. Available at: https://cointelegraph.com/news/nft-trader-hacked-millions-dollars-nft-stolen (accessed 30 August 2024).

Pfotenhauer S and Jasanoff S (2017) Panacea or diagnosis? Imaginaries of innovation and the 'MIT model' in three political cultures. *Social Studies of Science* 47(6): 783-810. https://doi.org/10.1177/0306312717706110

Pinto-Gutiérrez C, Gaitán S, Jaramillo D, et al. (2022) The NFT Hype: What Draws Attention to Non-Fungible Tokens? *Mathematics.* 10(3): 335. https://doi.org/10.3390/math10030335.

Rizzo J (2022) A Bored Ape Lawsuit Won't Set the NFT Precedent Seth Green Wants. *Wired*. 26 May 2022. Available at: https://www.wired.com/story/seth-green-bored-ape-nft-stolen/ (accessed 15 January 2025).

Russell F (2022) NFTs and Value. *M/C Journal* 25(2). https://doi.org/10.5204/mcj.2863.

Schrader A (2024) *U.S. Charges Three British Nationals for 'Evolved Apes' NFT Scam*. Artnet. 18 June 2024. Available at: https://news.artnet.com/art-world/evolved-apes-nft-scam-2501708 (accessed 30 August 2024).

Schuelke-Leech B-A (2018) A model for understanding the orders of magnitude of disruptive technologies. *Technological Forecasting and Social Change* 129: 261-274. https://doi.org/10.1016/j.techfore.2017.09.033.

Scott WR (1995) *Institutions and organizations.* Thousand Oaks: Thousand Oaks : SAGE.

Segre F (2023) Greening NFTs: Tezos leads the way in sustainable blockchain technology for a greener NFT ecosystem. *Onemint*. 25 September 2023. Available at: https://blog.onemint.io/greening-nfts-tezos-leads-the-way-in-sustainable-blockchain-technology-for-a-greener-nft-ecosystem/ (accessed 15 January).

Sevilla G (2022) One year later: How has Facebooks's Meta pivot fared? *Emarketer.* 27 December 2022 Available at: https://www.emarketer.com/content/one-year-later-how-has-facebook-s-meta-pivot-fared (accessed 30 August 2024).

Sheoran N (2015) 'Stratified contraception': emergency contraceptive pills and women's differential experiences in contemporary India. *Medical Anthropology* 34: 243-258. https://doi.org/10.1080/01459740.2014.922081.

Smith G (2019) Can the Fed Kill Bitcoin? Navigating the Chokepoints of Tax Law and KYC. *Bitcoin.com News*. 10 August 2019. Available at: https://news.bitcoin.com/can-the-fed-kill-bitcoin-navigating-the-chokepoints-of-tax-law-and-kyc/ (accessed 15 January 2024).

Streeck W (2014) *How will capitalism end?* London: Verso.

Swartz L (2022) Theorizing the 2017 blockchain ICO bubble as a network scam. *New Media & Society.* 24(7): 1695-1713. https://doi.org/10.1177/14614448221099224.

Taherdoost H (2023) Smart Contracts in Blockchain Technology: A Critical Review. *Information* 14(2): 117. https://doi.org/10.3390/info14020117.

Tonelli E (2021) *Fidenza Artist Sells $7M Worth of Ethereum NFTs Buyers Haven't Seen Yet*. Decrypt. 25 October 2021. Available at: https://decrypt.co/84296/fidenza-artist-sells-7m-worth-of-ethereum-nfts-buyers-havent-seen-yet (accessed 30 August 2024).

Truby J, Brown RD, Dahdal A, et al. (2022) Blockchain, climate damage, and death: Policy interventions to reduce the carbon emissions, mortality, and net-zero implications of non-fungible tokens and Bitcoin. *Energy Research & Social Science* 88: 102499. https://doi.org/10.1016/j.erss.2022.102499

USAO (2023) Non-Fungible Token (NFT) Developer Charged in Multi-Million Dollar International Fraud Scheme. *United States Attorny's Office.* 5 January 2023.Available at: https://www.justice.gov/usao-edny/pr/non-fungible-token-nft-developer-charged-multi-million-dollar-international-fraud (accessed 15 January).

van der Maarel S, Verweij D, Kramer E-H, et al. (2023) "This Is Not What I Signed up for": Sociotechnical Imaginaries, Expectations, and Disillusionment in a Dutch Military Innovation Hub. *Science, Technology, & Human Values* 0(0): 01622439231211032. https://doi.org/10.1177/01622439231211032

van Lente H (2021) Imaginaries of innovation. In: Godin B, Gaglio G and Vinck D (eds) *Handbook on Alternative Theories of Innovation*. Cheltenham, UK: Edward Elgar Publishing, pp.23-37.

van Lente H, Spitters C and Peine A (2013) Comparing technological hype cycles: Towards a theory. *Technological Forecasting and Social Change* 80(8): 1615-1628. https://doi.org/10.1016/j.techfore.2012.12.004

Velkova J (2019) Data centres as impermanent infrastructures. *Culture Machine* 18. http://culturemachine.net/vol-18-the-nature-of-data-centers/data-centers-as-impermanent/.

Vitalik.eth (2018) To be clear, at this point I quite regret adopting the term "smart contracts". I should have called them something more boring and technical, perhaps something like "persistent scripts". *X (then Twitter)* 4:21AM 14 October 2018. Available at: https://x.com/VitalikButerin/status/1051160932699770882

Vitalik.eth (2022) The merge will reduce worldwide electricity consumption by 0.2%. *X (then Twitter)*. 4:30PM 15 September 2022. Available at: https://x.com/VitalikButerin/status/1570299062800510976

Wang Q, Li R, Wang Q, et al. (2021) Non-fungible token (NFT): Overview, evaluation, opportunities and challenges. *arXiv preprint arXiv:* https://doi.org/10.48550/arXiv.2105.07447.

Weber M (1978) *Economy and Society.* Berkeley, CA: University of California Press.

Whiddington R (2022) NFT Buyers Are Suing Justin Bieber, Madonna, and Bored Ape Yacht Club's Founders Over an Alleged 'Scheme' to Bilk Investors. *artnet*. 12 December 2022. Available at: https://news.artnet.com/art-world/nft-buyers-suing-yuga-labs-celebrity-scheme-2227757 (accessed 30 August 2024).

Wikipedia (2023) *You wouldn't steal a car*. Available at: https://en.wikipedia.org/wiki/You_Wouldn%27t_Steal_a_Car (accessed 15 January 2024).

Wilser J (2021) *'I'm Obsessed': Paris Hilton on NFTs, empowering female creators and the fuiture of art*. CoinDesk. 17 Apriol 2021. Available at: https://finance.yahoo.com/news/m-obsessed-paris-hilton-nfts-152041206.html (accessed 15 January 2025).

Wilson S (2022) Situating Conceptuality in Non-Fungible Token Art. *M/C Journal* 25(2). https://doi.org/10.5204/mcj.2887.

Yang M (2023) The vast majority of NFTs are now worthless, new report shows. *The Guardian*, 23 September 2023. Available at: https://www.theguardian.com/technology/2023/sep/22/nfts-worthless-price (accessed 15 January 2025).

Youshaei J (2022) How Paris Hilton Reinvented Her Career. *Forbes*. 4 February 2022, updated 9 July 2024. Available at: https://www.forbes.com/sites/jonyoushaei/2022/02/04/from-reality-tv-to-nfts-how-paris-hilton-is-re-writing-her-story/ (accessed 30 August 2024).

Zaslowsky D (2024) US SEC Surprises Crypto Community by Approving Ether ETFs. In: Baker McKenzie. 24 May 2024 Available at: https://blockchain.bakermckenzie.com/2024/05/24/us-sec-surprises-crypto-community-by-approving-ether-etfs/ (accessed 31 August 2024).

Zilko C (2022) Seth Green Show Based on His NFT Paused After NFT Is Allegedly Stolen. *IndiWire*. 30 May 2022. Available at: https://www.indiewire.com/features/general/seth-green-stolen-nft-1234729399/ (accessed 15 January 2025).

JDSR 7(1) (2025) 103-138

JOURNAL OF DIGITAL SOCIAL RESEARCH

10.33621/jdsr.v7i154885

Issue ownership in the online campaign for Dutch general elections

A topic modeling approach

Joren Vrancken[1], Tom Dobber[2] and Frederik Zuiderveen Borgesius[3]

[1] Independent, The Netherlands
[2] University of Amsterdam, The Netherlands
[3] Radboud University, The Netherlands

🖂 jorenvrancken@gmail.com

Abstract

Online political campaigns are often opaque, among other reasons because political parties often target their advertising to specific groups. Therefore, it is challenging for citizens, journalists, and academics to understand what political parties talk about in their campaigns, diminishing the public accountability of political parties. Through the lens of issue ownership theory, this study explores which issues Dutch political parties advertised on Meta during the 2021 national election. The study uses a relatively novel topic modeling process that is meant to limit human bias. We built a model that assigns issues to each ad (based on the ad text), creating a dataset of ad-issues matchings. The study is one of the first to present insights into the issues Dutch political parties communicated about during the national elections of 2021. Our findings show that issue owners are not the biggest advertisers on their issues and reveal that private gifts enable some political campaigns to claim ownership of many issues.

Keywords: political advertising; issue ownership; political communication; topic modeling

1. Introduction

Political campaigns use online advertising to communicate to prospective voters. Online advertising comes with several downsides such as privacy violations, potential for manipulation, and a lack of transparency (Zuiderveen Borgesius et al., 2018). Focusing on this latter downside, the opacity around online political advertising means that it is unclear to citizens, journalists, and academics which issues political parties talk about. Social platforms offer ad libraries, but these are limited (Leerssen et al., 2021; Leerssen et al., 2019). As a result, public accountability is slim, as campaigners could falsely present themselves as one-issue parties (Zuiderveen Borgesius et al., 2018), or make electoral promises to narrow electoral groups (Dobber & De Vreese, 2022). In this study, we aim to shed light onto this opaque realm of online political communication by examining the issues communicated through online political advertising by Dutch political parties in the national election campaign of 2021.

Research on online political advertising so far has focused on the effectiveness of online political advertising (e.g., Haenschen & Jennings, 2018; Haenschen, 2022; Zarouali et al., 2020; Endres, 2019; Lavigne, 2020; Coppock et al., 2020), on the circumstances that lead to the use of online advertising (Kreiss, 2016; Anstead, 2017; Dobber et al., 2017; Kruschinski & Haller, 2017), or on citizen perceptions of online political advertising (Turow et al., 2012; Dobber et al., 2018). Few studies have focused on the content of online political advertisements and those that did (Kruikemeier et al., 2022; Fowler et al., 2020) either focused on the United States, or had a narrow scope (Dobber & De Vreese, 2022). This study takes a broad approach, focusing on all parties and all advertisements on Meta that include a policy element, and focuses on the Netherlands, a European multiparty context. The Netherlands is a somewhat extreme example of a country with a multi-party system, as 17 parties gained at least one seat in parliament after the 2021 elections. Hence, the Netherlands stands in clear contrast to the often-studied United States, with two leading parties.

Political parties benefit from online advertising because it affords political campaigns more control over their message, and less reliant on the agenda-setting news coverage of the traditional media (McCombs & Shaw, 1972). Online advertising can be understood as an alternative communication channel, next to the traditional media, that can help political campaigns reach potential voters on the issues that they own. Issue ownership theory (Petrocik et al., 2003) suggests that "some political parties are affiliated with specific issues and considered best able to deal with them" (Walgrave et al., 2015, p. 778). In other words, voters see certain problems (or issues) as a typical focus point for certain parties, and may think that those parties can best address those problems. Reaching out to potential voters on owned issues seems beneficial for parties, as this strategy could improve party support (Endres, 2019 Hillygus & Shields, 2008; Abbe et al., 2003) and might increase the vote share among volatile voters (Geers & Bos, 2017). US-focused research has indeed shown that political campaigns focus on their owned issues in their online campaigns (Kruikemeier et al., 2022).

However, Zuiderveen Borgesius et al. (2018) warn that the opacity of online advertising could enable political parties to falsely present themselves as one-issue parties to different voters. In other words, through online advertising, political parties could emphasize issue A to voters who are thought to find issue A important, and then repeat this process for issue B, and for voters who care about issue B. This makes it difficult for voters to understand which issues are important to political parties and which issues are less important.

Catering to different voter groups by emphasizing different types of issues also makes it more difficult for political parties to interpret their mandate once elected (Jamieson, 2013). For example, when a party campaigns solely on an immigration platform, it is easy for the voters to understand the policy priorities of that party. Once elected, politicians from that party have a clear mandate. But when that party campaigns on immigration, education, environmentalism, security, fiscal responsibility, and ten other different issues, the priorities and mandate of that party are less clear-cut.

The few studies that focused on the content of online political advertisements examined the United States (Kruikemeier et al., 2022; Fowler et al., 2020). However, there are clear contextual differences between de facto two-party systems such as the US and multiparty systems that are often found in Europe, especially through the lens of issue ownership theory (Petrocik et al., 2003). Most importantly, in the US, only two parties divide and contest the issues among each other, but in European multiparty systems the issues are divided and contested among many more political parties. Especially in the Netherlands, where since 2021 the national parliament counts 17 political parties (Kiesraad, 2021), the issues are much more contested (see Appendix A for an overview of Dutch political parties).

Analyzing the issues communicated in online political campaigns is crucial to understanding parties' policy positions. However, identifying these issues is challenging because online ads come in many forms that range from just a few words to multi-paragraph articles. Often, researchers use topic modeling to identify topics on a large scale; however, such techniques often "require the additional step of attaching meaningful labels to estimated topics". Therefore, topic modeling is sometimes critiqued for the human bias it introduces (Béchara et al., 2021, p. 1). This current study builds upon Béchara et al. (2021) and Kruikemeier et al. (2022) and assigns issues to political ads using a pre-defined list of words that are relevant to an issue based on expert codebooks. Through the lens of issue ownership theory, this study explores which issues Dutch political parties advertised on Meta during the 2021 national election certain issues. In doing so, the study applies a relatively novel topic modeling process to limit the human bias often found in topic modeling.

The paper combines insights from different fields, including communication science, political science, and methods from computer science.

2. Theoretical framework

Issue proximity theorists (e.g., Downs, 1957) argue that citizens vote for the parties or politicians that most closely resemble their own issue positions. However, it becomes increasingly challenging for citizens to compare their own issue positions with those of political parties. Although political parties generally publish manifestos on their websites, manifestos are often difficult to comprehend for citizens (Bischof & Senninger, 2018). Therefore, few citizens read manifestos (Adams et al., 2014; Adams et al., 2011; Andersen et al., 2005). Encountering information about issue positions in the mainstream media is challenging because the mainstream media tend to cover the larger parties (Kostadinova, 2017), and focus on horse race, conflict and campaign strategy news (Ergün & Karsten, 2019).

Online advertising techniques might enable political campaigns to communicate directly to the electorate, without interference from critical journalists, but these advertising techniques also enable campaigns to target and tailor their ads to the issue preferences of the targeted subsegments of the electorate. Since the online advertising infrastructure is opaque, it is unclear to the citizen to what extent a political party genuinely prioritizes a specific issue, or whether that party pretends to prioritize that issue because data analysis reveals that the targeted audience cares for that issue (see Zuiderveen Borgesius et al., 2018).

According to issue ownership theory "some political parties are affiliated with specific issues and considered best able to deal with them" (Walgrave et al., 2015, p. 778). Political parties campaign use issue ownership cues and issue position cues (Banda, 2016). An issue ownership cue can be a Green Party advertisement about how the environment is in good hands with them, and an issue position cue signals to the citizen what their stance is within that issue: the Green Party suggesting closing all coal-fueled power plants, for instance. In the Netherlands, the largest party VVD ran a campaign in 2012 stating that 'the economy could use some VVD', which is a clear issue ownership cue.

Issues ownership can change over time, and issue ownership can also be contested. This occurs when citizens do not clearly perceive one party most competent to handle a specific issue (Geys, 2012). It rarely happens that a party has complete issue ownership and over time, parties might emphasize certain issues more than they did before. The environment, for instance, used to be an issue clearly owned by green party GroenLinks, but over time this issue is likely to be also emphasized and claimed by other types of political parties (i.e., issue trespassing (e.g., Walgrave et al, 2009; Bos et al., 2016). Indeed, Walgrave et al. (2009) found that issue ownership is subject to change and can be contested through news coverage. Scholars do not agree about when a party 'owns' an issue. Petrocik (1996) employs a 50% threshold, but this is in a US context. In a 17-party democracy such as the Netherlands, it is unlikely that any party is seen by over 50% of the electorate as most competent to handle any issue. Walgrave and De Swert (2017) identified strong issue ownership, which occurs when one party is seen by around 50% of the electorate as most competent, and intermediate issue ownership, which occurs when ownership is shared between parties or with one "slightly dominating party" (p. 43).

Political parties can also claim previously unclaimed issues through the news cycle. Claiming issues and running campaigns on owned issues can be a useful strategy. In a rationalistic conception of voting behavior, people are expected to vote for parties that 'own' the issue found most salient (e.g., Downs, 1957). Some empirical evidence suggests that contacting or cross-pressuring citizens on a (personally) salient or owned issue increases candidate or party support (Endres, 2019 Hillygus & Shields, 2008; Abbe et al., 2003; Walgrave & De Swert, 2010). Research by Geers and Bos (2017) shows that volatile voters are more likely to vote for issue owners, especially when these parties are visible in the media and covered positively.

On the other hand, there is a risk in campaigning on unimportant or niche issues. Reaching out to a large and heterogeneous group with ads about owned issues that are perceived too niche can be counterproductive. For instance, the Orthodox Calvinist Christian party in the Netherlands might campaign to bring back the Christian oath for public servants. However, when the majority of potential voters cares about abortion, the former issue-ad might backfire, because voters might think that the party neglects their most salient issue (similar to Chou & Lien, 2010).

This study aims to extend issue ownership theory (Petrocik, 1996), by applying it to the online advertising context. The affordances of online advertising enable campaigns to run highly differentiated campaigns. For example, campaigns can reach out to a certain subset of voters for which issue a is deemed most salient, and simultaneously reach out to voters for which issues b or c are most salient. This is different from traditional advertising, which does not afford granular opportunities to differentiate messages among audiences. As a result, traditional advertising would need to rely much more on the salient issues than online advertising would. In other words: through online advertising, campaigns or no longer confined to the salient issues and they can advertise on many more issues compared to traditional advertising. This is especially pressing in light of the issue competition between parties in Europe in general (Green-Pedersen, 2007), but particularly in the politically fragmented Dutch context. Especially in the Netherlands, with many competing parties, parties can use online advertising to talk about more different issues. Where parties using traditional channels are incentivized to focus on the most salient issues, in online advertising they do not need to focus on salience perceptions alone because they can diversify their messages among different audiences' issue preferences. However, this could happen in such an opaque way that it is challenging to get a comprehensive overview of a party's viewpoints (Zuiderveen Borgesius et al., 2018). This study attempts to provide a comprehensive overview, starting by answering the following research questions.

RQ1.On how many different issues does each party communicate, and what is the total number of impressions per issue per party?

2.1 Issues owned versus issues claimed

After the national election of 2021, the Dutch parliament counted 17 political parties. Based on an opinion poll conducted one month before the election, not one party was an undisputed issue owner (I&O Research, 2021). However, when we focus on parties that were perceived as issue owners by a majority of the people, we see that only two parties are issue owners.1 The largest party in the Netherlands, the VVD, owns the issue 'economy' (according to 58% of the respondents). The PVV owns the issue 'migration' (59%).

Van der Meer and Damstra (2022) measured associative issue ownership. They asked people which party they associated with specific issues, regardless of perceived competence. Hence, they did not ask which party people deemed most competent to address a certain issue, as was the case in the opinion poll of I&O Research (2021). Van der Meer and Damstra (2022) found that VVD owns economy (according to 70% of respondents), the PVV owns migration (55%), and GroenLinks owns climate change (60%). However, prominent issues such as housing are much less unequivocally owned (Van der Meer &

[1] Provided that we use the golden standard CAP issue list. If we use the issue list provided by the pollster, we see that the VVD also owns 'government finance' (53%) and the Animal Party owns 'animal welfare' (80%).

Damstra, 2022), which is in line with the opinion polls that measured issue ownership based on competence perceptions (I&O Research, 2021).

Since issue ownership perceptions have been shown beneficial for voter support and even voter behavior, especially when combined with visibility (Geers & Bos, 2017), one can expect issue owners to communicate strongly on 'their' issues. However, many issues remain challenged (as identified by the Comparative Agendas Project). Moreover, since the 2021 election saw three new political parties arise (BBB, Volt, and JA21), there is ample opportunity to reshuffle the board.

This leads to the following hypothesis and research question.

H1a. Compared to other topics in VVD ads, the VVD spends the most money on ads on the economy and gets the most impressions on ads about the economy.

H1b. Compared to other parties, the VVD spends the most money on ads on the economy and gets the most impressions on ads about the economy.

H2a. Compared to the other topics in GroenLinks ads, GroenLinks spends the most money on ads on the environment, and gets the most impressions on ads about the environment.

H2b. Compared to other parties, GroenLinks spends the most money on ads on the environment and gets the most impressions on ads about the environment.[2]

RQ2. Which parties claim which issues in terms of ad spending and number of impressions?

2.2 Issue ownership per consideration set

Citizens in multiparty systems do not consider all political parties when determining their vote choice. Rather, undecided citizens have a consideration set of potential political parties they consider voting for. Citizens may not vote for the same party each election, but they are likely to vote for a party within their consideration sets (Rekker & Rosema, 2019).

According to panel survey research done in the Netherlands (Rekker & Rosema, 2019), there is a leftist camp of parties that consists of SP, GroenLinks, PvdA and D66. There is a Christian camp consisting of CDA, ChristenUnie and SGP. There is a rightist camp (VVD, D66, CDA). And a radical right camp that, in 2019, consisted of the PVV, but would now likely also include FvD and potentially JA21. D66 is placed in both the rightist camp and the leftist camp: this is because D66 is a center party and consideration sets are based on citizen perceptions.

Because citizens are unlikely to vote for a party outside of their consideration set (Rekker & Rosema, 2019), political parties are unlikely to target their advertisements to people who hold different consideration sets. For example, rightist parties are unlikely to target leftist voters and vice versa. This would suggest that we should not only examine issue ownership on the scale of all political parties, but also that we should take the consideration sets into account. Since a leftist party does not target rightist voters, the leftist party could claim issue ownership within the consideration set. As the rightist party VVD is considered the overall issue owner on economy, it is unlikely that the other parties in the rightist camp will run ads on the economy because the VVD will target the same voters and do this more convincingly when it comes to the economy. But the VVD is less likely to target voters in the leftist camp, leaving a vacuum in which the economy can be claimed by a leftist party within that consideration set. This leads to the following research question:

RQ3.Which parties claim which issues in terms of number of ads and number of impressions within consideration sets?

[2] We know in advance that the PVV has placed only one issue advertisement on Meta's platform, so we did not formulate a hypothesis for the PVV.

3. Method

In this study, we analyze ads that ran on Meta (i.e. Facebook and Instagram) during the political campaigns leading up to the 2021 elections. We built a model that assigns issues to each ad (based on the ad text), creating a dataset of ad-issues matchings. This dataset allows us to aggregate data from ads about specific issues.

3.1 The Meta ad library

The Meta Ad Library, released in 2019, provides all (political) ads published on Facebook and Instagram. We chose to focus on the Meta Ad Library as it provides more detailed information on the content and targeting of ads than other social media ad repositories (e.g. Google's Ads Transparency Center or the TikTok Ad Library). This limits our research to ads that ran on Meta's platforms. However, we do not consider this a significant limitation, as most political parties heavily advertise on Facebook and Instagram.

Table 1. The number of Facebook ads per party

Party	**# of Ads**
CDA	11,463
VVD	4,461
PvdA	4,425
D66	1,527
VOLT	1,180
GL	444
SP	424
DENK	414
FvD	283
PvdD	276
CU	228
BIJ1	186
50+	162
JA21	115
SGP	104
BBB	60
PVV[3]	9
Total	**25,761**

[3] As the PVV is a controversial party, they get ample natural online attention from both proponents and opponents. This might explain why the PVV spends little on online advertising.

https://doi.org/10.33621/jdsr.v7i154885

A Meta ad consists of three elements: a main body of text, an image or video and a call to action (e.g., a link to a website of a political party). Not all these elements must be present (e.g., an ad can have a body but no image).

The Meta Ad Library provides an API for programmatic extraction of ad data. For each ad, Meta provides its content (i.e. the body, image and call to action) and metadata about the demographics of the audience the ad was shown to (i.e. gender, age bracket and geographical region) and statistics on how the ad performed (i.e. impressions, potential audience and spending). In this study, we focus on the content of the ads.

The large disparity between the number of ads between parties reflects the structure and digital marketing strategy of the parties. Parties that have many local branches tend to run more ads, as each local branch runs its own ads. In general, this means that ads by parties with local branches get lower impressions than ads by other parties, as the local branch ads are meant for smaller audiences. It does not mean that parties with fewer ads get fewer impressions overall.

Some parties prefer to run ads with larger texts that cover multiple issues and talking points, while other parties prefer multiple smaller ads that each cover a single issue.

We limit the impact of this disparity on the results by designing the model for ads with larger bodies (covering multiple issues) and smaller bodies (covering only a single issue).

3.2 Issues

The Comparative Agendas Project (n.d.) is a research project that provides codebooks for issues (and sub-issues) that cover the broad public debate. Following the issues, listed in the Netherlands-specific codebooks, we created the following list of 14 issues that cover a broad spectrum of the Dutch political debate:

Agriculture
Civil Rights
Climate
Defense
Economy
Education & Culture
Foreign Affairs
Government
Healthcare
Housing
Law & Order
Migration
Social Welfare
Transportation

As not all issues provided by the codebook are distinct from one another (in the context of Dutch political campaigns), we merged similar issues (e.g., international affairs and foreign trade).

3.3 Matching ads and issues

To match an ad to one or multiple issues, we compare the text[4] in the ad (both the body of text and text provided in the call to action) to each issue word list, by computing the intersection between the ad text and each issue word list. We consider an ad to be about an issue if one of the following conditions is true:

1. The issue word list has the largest intersection with the ad text and the cardinality of the intersection is larger than one.
2. The cardinality of the intersection between the issue word list and the ad text is larger than five.

Consequently, an ad can be matched to no issues, one issue or multiple issues. We allow for this flexibility to accommodate as many types of ads as possible, as ad texts are not uniform and range from a few words about a single issue to multi-paragraph texts covering a broad range of issues.

We give an example of this approach based on two simplified issue word lists (Table 2) and three example ads:

[4] Specifically, the lemmatized forms of the nouns, proper nouns and adjectives in texts.

Table 2. Two example issue word lists for the housing and climate issues

Housing	Climate
House	Sustainable
Starter	Green
Building	Climate
City	Energy
Mortgage	Windmill
Rent	Environment
Housing	Clean
Residence	Solar

Example ad 1: “There are far too few houses in the Netherlands, especially for starters. That is why we are going to build 1 million new houses together.”

Example ad 1 has the largest intersection with the issue word list of housing, because the ad text has three words in common with the housing issue word list and zero with the climate issue word list. As such, is mapped to the issue housing.

Example ad 2: “Come to our congress on the 2nd of October! Buy your tickets now.”

Example ad 2 does not discuss a specific issue and consequently does not have any words in common with the issue word lists. It is not mapped to either issue.

Example ad 3: “We will invest heavily in the development of new sustainable housing. Future-oriented projects, sustainable alternatives and new green technologies, we ensure that they can be set up and developed in the regions. We focus on windmills and solar panels, to ensure a clean future with clean energy in clean cities. This will ensure cheaper rents and lower mortgage rates, especially for starters looking to buy their first house.”

Example ad 3 has the largest intersection with the issue word list of climate (eight words), meaning it will be mapped to the issue climate. However, because it also has six words in common with the issue word list of housing, it is also mapped to housing.

As the Comparative Agendas Codebook codebooks are not tailored to a specific election, they contain words that are not relevant to the 2021 election and miss words that are. We solved this by manually updating the issue word lists with common words from the ads matched to each issue. We iterated this process until we could not find any new relevant words in the matched ads. The final word lists can be found in Appendix C.

In total, we matched 11,336 ads to at least one issue. If an ad was matched to at least one issue, we consider it “matched”. If an ad was not matched to at least one issue, we consider it “unmatched”. Appendix F shows the distribution of matched and unmatched ads per party. As we can see, a large percentage of ads are unmatched. This is expected, as political parties do not only run ads about political issues, but also run ads about organizational matters (e.g. “come to our party congress next month”) and ads about party representatives (e.g. “Please meet the candidate representative from your city.”). Some ads were not matched to any issue, because they contain too little text for the model to match the ad to a specific issue with a level of certainty.

Appendix G shows the distribution of issues per party and the total of ads per issue.

To validate the dataset created by our model, we computed an inter-coder reliability measure of a human coder and our model for a random sample of 300 ads. One human coder manually went over each ad in the sample and noted which issues they considered the ad to be about (either zero, one or multiple issues). We compared the manual encodings with the encodings of our model for the same sample of ads, by computing the Krippendorff's alpha.

Because ads can have multiple issues, the coders can have a partial agreement (e.g. they agree that an ad is about an issue but disagree that the ad is about another issue). To take partial agreements into account, we computed the multi-label Krippendorff's alpha, using the MASI distance function (Passonneau et al., 2006). This resulted in a score of .84. Scores above .8 are generally considered reliable enough for meaningful interpretation (Krippendorff, 2004).

In appendix D, we provide the Krippendorff's alpha for each individual issue.

3.4 Analysis

Besides text, the Meta Ad Library also provides statistics on each ad. The most important statistics that the Ad Library provides are money spent, impressions (i.e., the number of times an ad was shown) and potential reach (i.e., how many users could have seen the ad). Meta does not give these statistics as absolute numbers, but as ranges (e.g., between €2000, - and €3000, - was spent on an ad). We aggregated the data on ads about each issue to analyze the ad data on an issue-level (e.g., the number of impressions for agriculture).

4. Results

4.1 Issues and impressions per party

The first research question asked: on how many different issues does each party communicate, and what is the total number of impressions per party? Appendix B shows that center-right party CDA is an outlier in terms of total issue-related ad impressions (over 46 million impressions). By contrast, the issue-campaign of the biggest party in terms of parliamentary seats VVD received over 10.5 million impressions. This latter number is more in line with the other 'larger' parties that cater to a larger and more heterogeneous section of the electorate. The smaller parties such as Calvinist party SGP (over 600.000 impressions) and immigrant party DENK (over 1.2 million impressions) in general received fewer impressions than the larger parties. Another outlier is the PVV, which is the third party of the Netherlands in terms of size. The PVV only received 2.500 impressions and only ran one ad on the issue climate. In terms of issue diversity of the online campaign, Green Party GroenLinks was the only party that campaigned on all fourteen issues. Only five parties campaigned on the issue foreign affairs, making this the issue contested by the least number of parties. Climate, housing and healthcare were campaigned on by fifteen parties (out of 17). Appendix E shows the spending per issue, per party.

4.2 Issue ownership VVD

The first hypothesis expected that VVD spends the most on ads on the economy and gets the most impressions on ads about the economy in comparison with a) the other VVD ads and compared with b) the other political parties. Focusing first on the first part of H1, Figure 1 shows that VVD runs the most ads on housing (34%), followed by economy (14%), healthcare (13%), and climate (11%). Figure 1 shows that VVD also spent most on issue-ads relating to housing (EU 22,000), followed by climate (EU 9,500), and healthcare (EU 7,500). VVD spent slightly under 4,500 euro on ads about the economy. In terms of impressions, housing ads make up for 24% of total impressions and economy ads receive 5% of the total impressions, even though 14% of ads are about the economy. This means that hypothesis 1a is not supported. In fact, the VVD runs the most ads on housing, spent the most on housing ads and also receives the highest number of impressions on housing-related ads.

Figure 1 shows, for instance, that the VVD spends a lot of money on ads about healthcare (third place in the list at the left), but the VVD ads on civil rights receive more impressions. The difference between

https://doi.org/10.33621/jdsr.v7i154885

the ranking in “money spent per issue by VVD” and “impressions per issue by VVD”, could have several reasons. For instance, some more expensive ads could be less engaging, or targeted at more expensive audiences. Unfortunately, the ad delivery algorithm is opaque, and a definitive explanation is beyond the scope of this study.

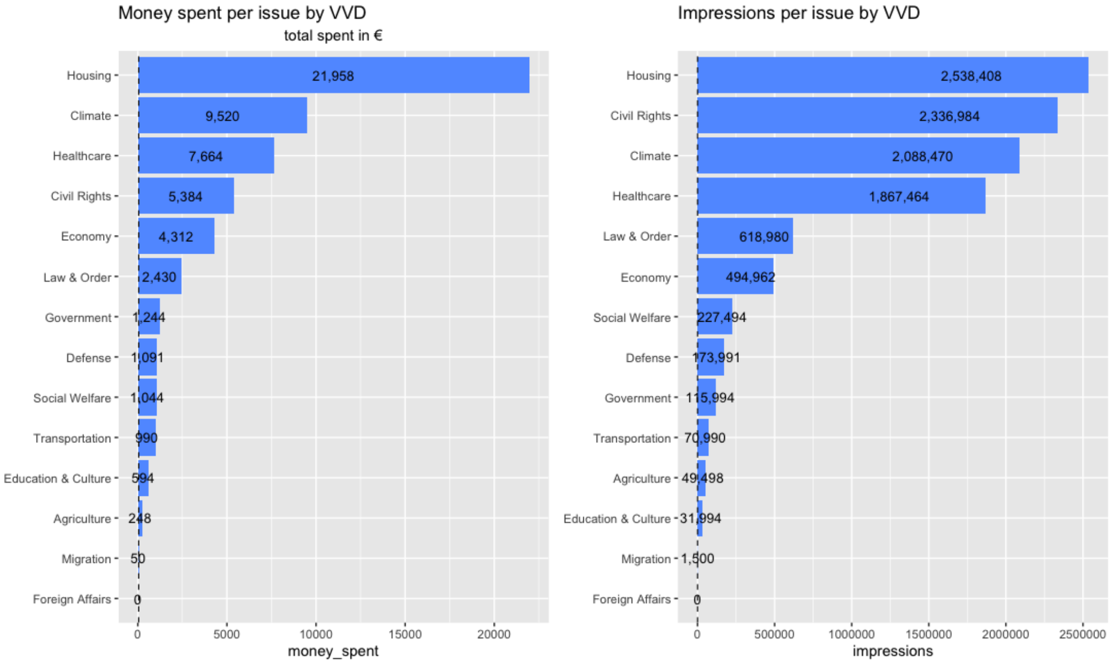


Figure 1. Money spent (in EUR), and impressions gained per issue

Moving on to the second part of H1, figure 2 shows that VVD does not pay the most for ads related to the economy and does not gain the most impressions on these ads. This means that H1b is not supported. In terms of the number of ads about the economy, it is evident that CDA dominates the online space. Of all economy-related ads, 73% are from the CDA. CDA also paid over 89,000 euro for economy ads (compared to slightly under 4,500 euro for VVD). In terms of impressions, however, CDA is less dominant: of all the impressions related to economy ads, 39% are for CDA ads. The economy ads of (radical) rightist parties JA21 and FvD perform strongly: both parties spend less than 1% of the total amount spent on economy-related ads, but both receive 10% of the total number of economy-related ad impressions.

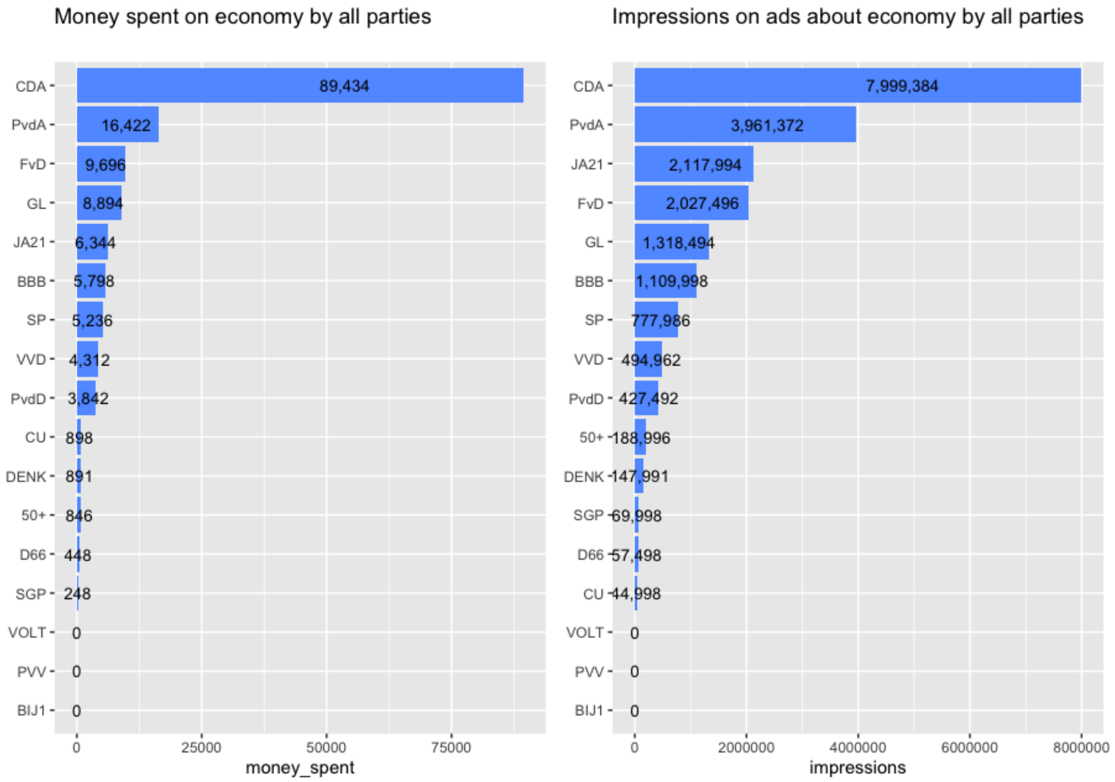


Figure 2. Money spent (in EUR) and impressions gained by ads about the economy

4.3 Issue ownership GroenLinks

The second hypothesis expected a) that GroenLinks, as issue owner on climate, would spend the most on climate ads, and b) would get the most impressions relative to the other GroenLinks issue ads on Meta. Figure 3 shows that, in terms of spending, this is clearly the case: GroenLinks spent 79,390 euro on issue ads. Almost 24,500 euro, so 31% of all the money that GroenLinks spent on issue ads went to ads about climate change, followed by healthcare (€10,290 or 13%). The data paints a similar picture in terms of impressions. GroenLinks' issue ads were displayed just over 13 million times (13,099,890). Almost 4 million impressions went to climate ads (30%), followed by healthcare (slightly over 2 million impressions, 16.5%).

The second hypothesis also expected that GroenLinks would spend more on climate change ads and get more impressions than each individual other party. Figure 4 shows that this is not the case. PvdD is the biggest spender (€ 54,943), followed by GroenLinks € 24,467). In terms of impressions, PvdD (5,534,443 impressions) also trumps Groenlinks (3.965.967).

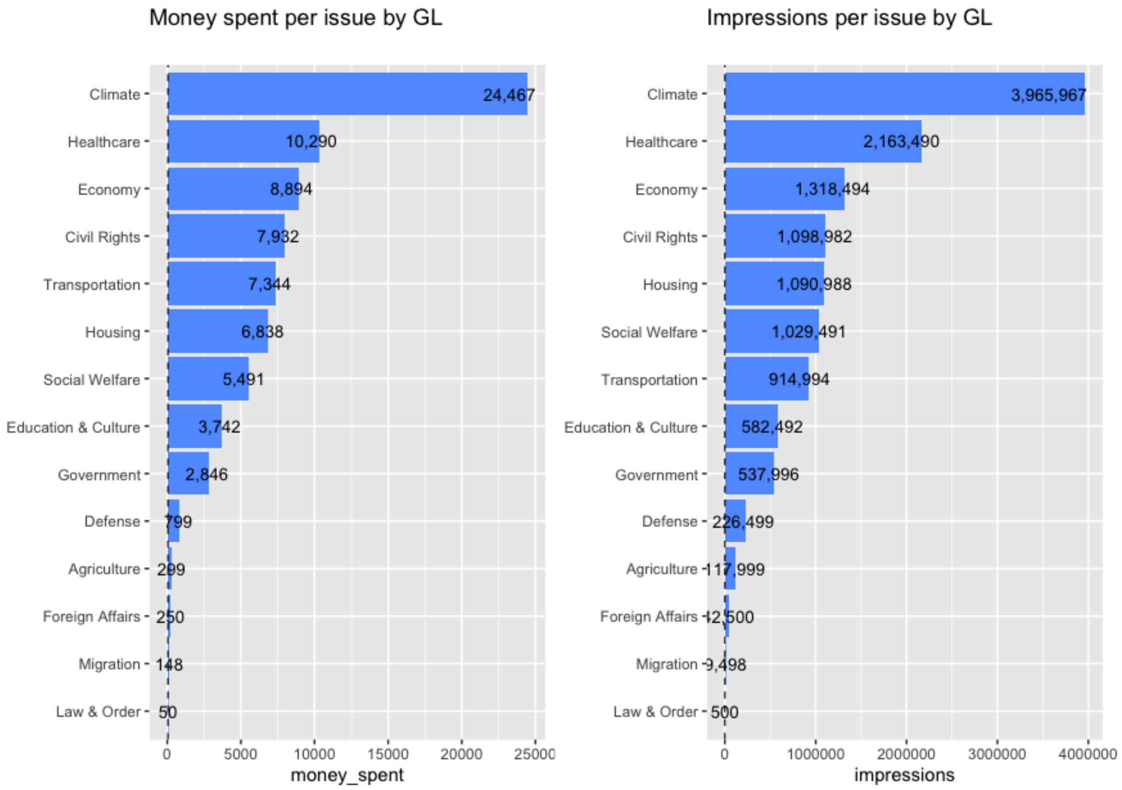


Figure 3. Money spent (in EUR) and impressions gained per issue GroenLinks

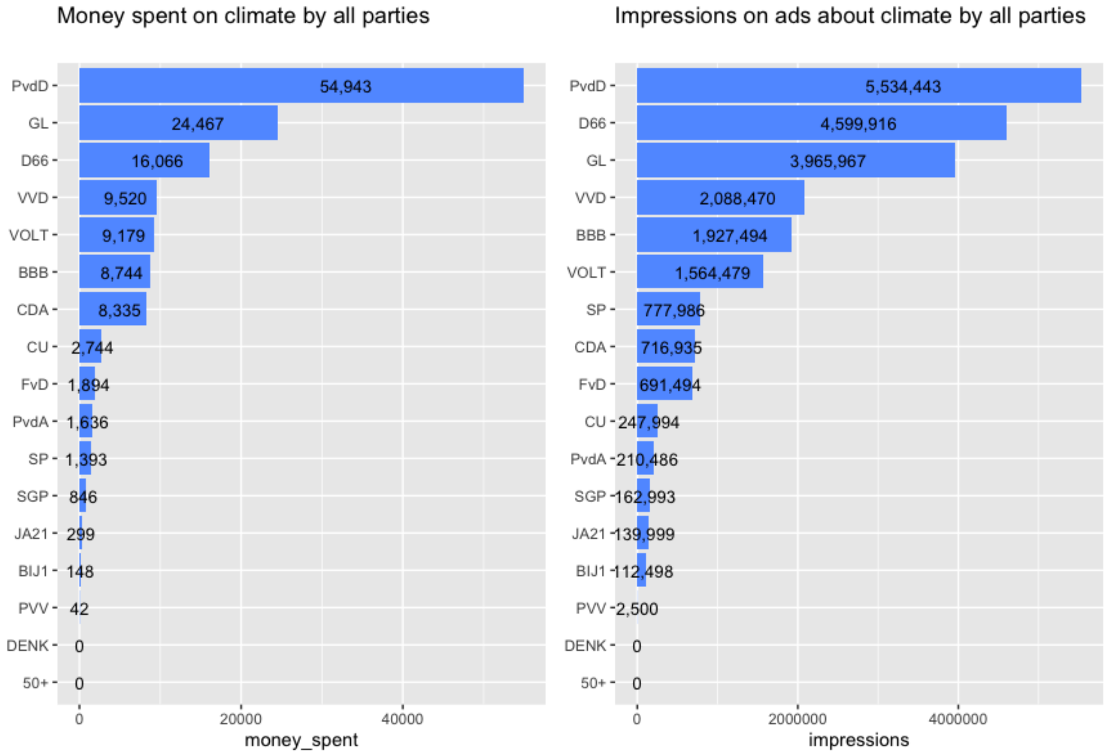


Figure 4. Money spent (in EUR) and impressions gained by ads about the climate

4.4 Which parties claim which issues?

Research question 2 asked which parties claim which issues in terms of ad spending and number of impressions. Figure 5 shows that major party VVD does not claim any issues in terms of advertising spending. In fact, center-right party CDA is the biggest claimer of issues in terms of spending. CDA spends more than the other parties on civil rights, economy, government, housing, and social welfare. Remarkably, CDA used to be known as the 'farmer's party', but CDA was challenged strongly by BBB and PvdD on the issue agriculture in terms of money spent and impressions reached (see Appendix B). The most strongly contested issues in terms of money spent and impressions reached are civil rights (claimed strongest by CDA), social welfare (CDA), housing (CDA), and economy (CDA). Surprisingly, climate was claimed strongest by PvdD and not GroenLinks in terms of money spent and impressions reached. Remarkably, while PvdA spent more than the other parties on education-related ads, D66 reached the most impressions with their education-related ads. Similarly, while PvdA spent more than other parties on healthcare ads, anti-lockdown party FvD reached the most impressions with their healthcare ads.

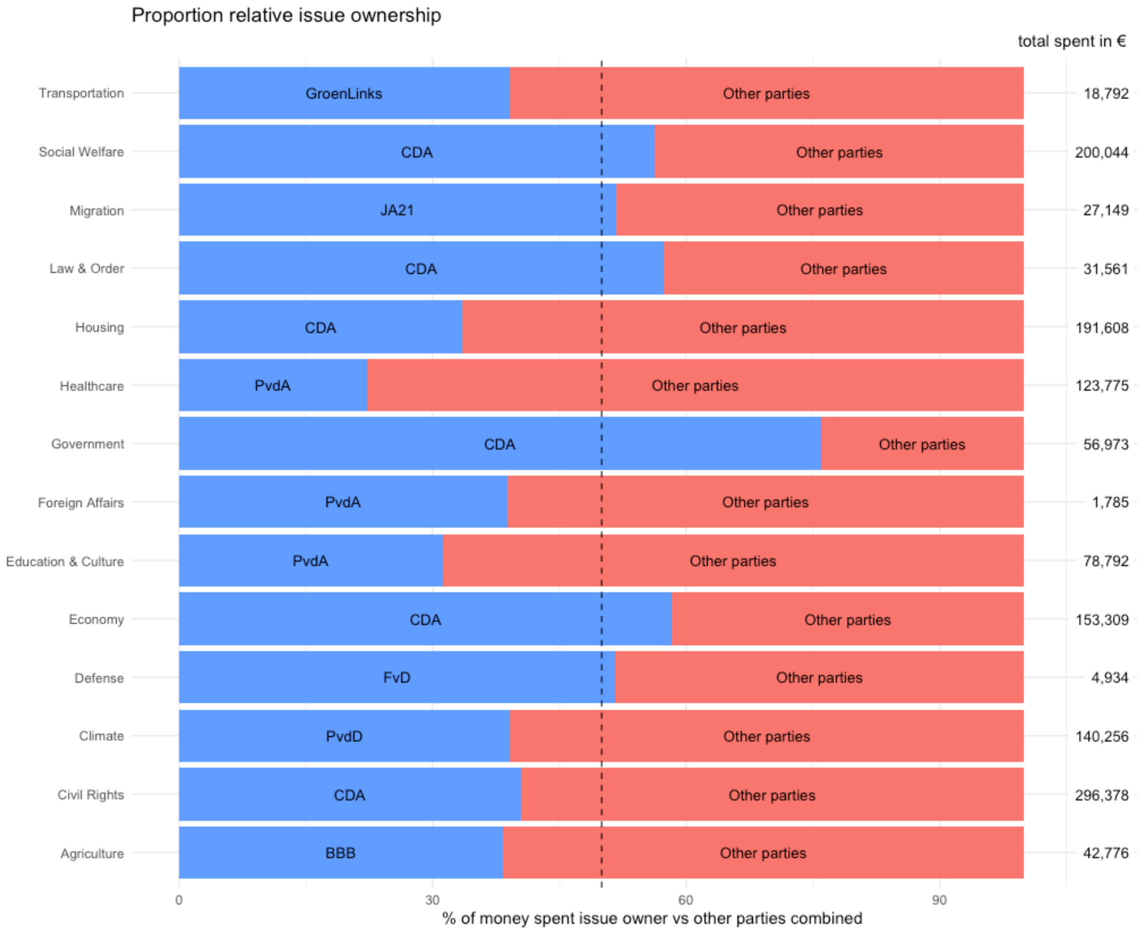


Figure 5. Issue claims in terms of money spent (in EUR) for all parties

Note. This figure displays alphabetically the degree to which each issue is owned by a specific party. In other words: the biggest spender is shown on the left, in blue, and the amount spent by this party is contrasted by the combined spending of the remaining parties (on the right). The dashed line denotes 50%, which means that the biggest spending was 50% of the total amount spent on ads about a specific issue. The further the left-hand bar passes the dashed line, the stronger that party has claimed the issue in terms of online advertising.

https://doi.org/10.33621/jdsr.v7i154885

4.5 Which parties claim which issues within consideration sets?

We distinguish four consideration sets: the leftist parties (D66, GroenLinks, PvdA, SP, Bij1, PvdD, Volt, Denk), Christian parties (CDA, ChristenUnie, SGP), rightist parties (D66, CDA, VVD, BBB, Volt), and radical right parties (PVV, FvD, JA21). We compare claims on issue ownership for each consideration set. For the leftist consideration set we find that climate and civil rights are the most contested issues. For climate PvdD claims this issue by spending slightly more than 50% of the total sum spent on this issue by all leftist parties. Civil rights are claimed strongest by PvdA, by spending 55% of the total sum spent in the leftist consideration set. PvdA did not strongly claim issues when we look at all parties combined, but in the leftist consideration set, the PvdA is the dominant party with regard to spending (see Figure 6). In the Christian consideration set, CDA dominated strongly on all issues in terms of spending. For the rightist consideration set, CDA was the biggest spender on all issues except agriculture (BBB), climate (D66) and defense (VVD). See Figure 6. In the radical right consideration set, FvD claimed most issues. But on key radical right issues migration and law and order, JA21 held the strongest claim in terms of ad spending (see Figure 6).

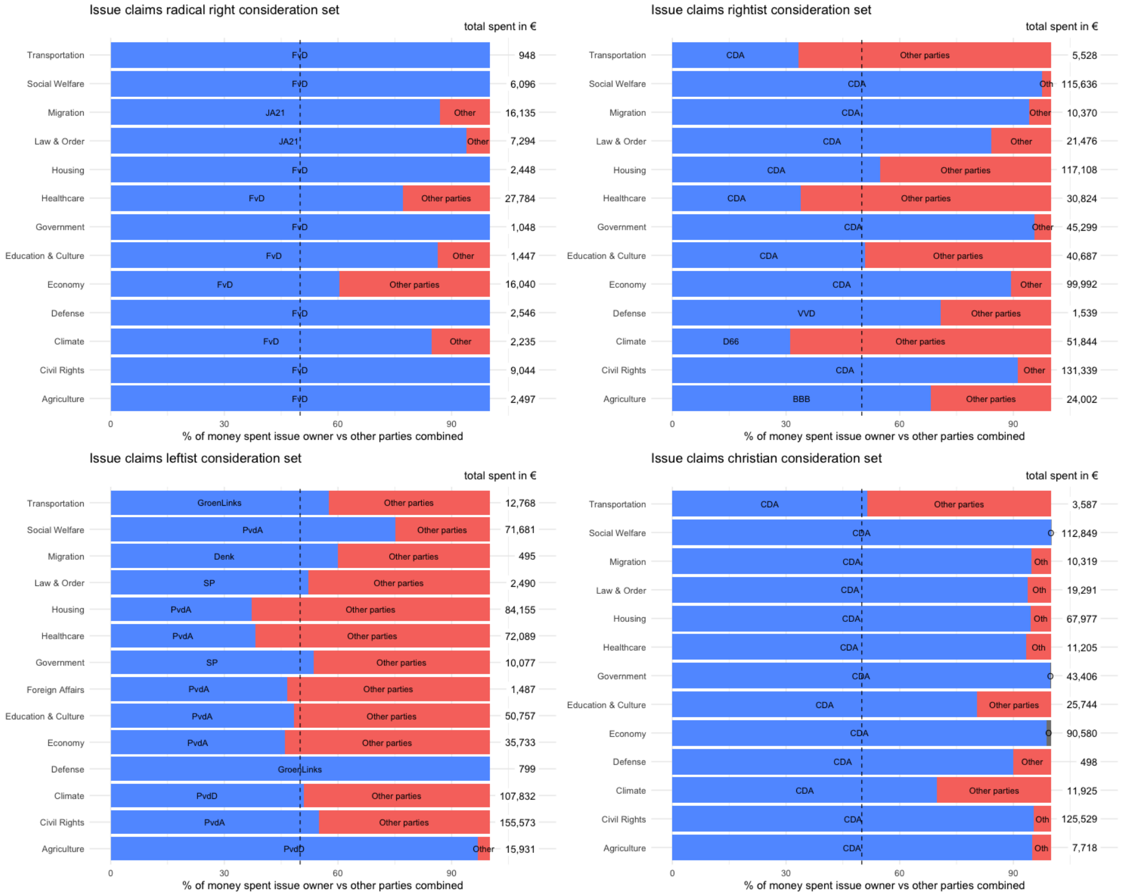


Figure 6. Issue claims in terms of money spent (in EUR) per consideration set

5. Discussion

In this study we explored which issues Dutch political parties advertised on Meta in the run-up to the 2021 national election. Interpreting the findings through the lens of issue ownership theory (Petrocik, 1996), we see that the affordances of online advertising indeed drive issue trespassing as well as issue competition. The owned issue economy (owned by VVD) was claimed by CDA, who spent most money

https://doi.org/10.33621/jdsr.v7i154885

on this issue and reached most impressions. Similarly, the owned issue climate (owned by GroenLinks) was claimed by Partij voor de Dieren, who spent most money on this issue and reached most impressions.

In terms of competition, we see that political parties do not limit their communication to only a few salient issues since each issue was contested by a number of parties (see Figure 5). Hence, this study suggests that political parties use online advertising to appeal to voters in a differentiated way. Where traditional advertising confines parties to only the salient issues, online advertising affords parties the opportunity to reach out to voters on less salient issues.

We found that social welfare, housing, healthcare, economy and climate were the most hotly debated issues in the online campaign on Meta. Since the data pertain to the 2021 election, when covid19 was still prominent, it is not surprising the health care was among the top-debated issues in this campaign. However, in the 2023 election campaign, healthcare is still a salient issue in the minds of the aging Dutch population (Kanne & Van de Koppel, 2023).

We approached this study through the lens of issue ownership theory and found that the issue owner regarding economy, the VVD, was strongly outspent by CDA and 7 other parties. The issue owner of climate, GroenLinks, was strongly outspent by small party PvdD. This is striking because when Dutch citizens were asked about important societal problems, there was most agreement about climate change being an important problem (Van der Meer & Damstra, 2022).

The fact that both issue owners were outspent on their owned issues can be partly explained by the impact of individual benefactors on the relatively underfunded Dutch election campaigns. In the run-up to the 2021 campaign, CDA received a gift of 1 million euro from one entrepreneur (NOS, 2021). Similarly, D66 received 1 million euro and PvdD received 350,000 euro as a gift from an entrepreneur (Parool, 2021). Dutch election campaigns are underfunded in comparison with similar European campaigns (Andeweg et al., 2008). For example, the largest party VVD spent 2,718,325 euro on their entire campaign for the 2021 national election (NPO Radio 1, 2021). Therefore, such large gifts can tilt the playing field toward receiving parties. Parties reassure that benefactors do not buy political influence (e.g., NOS, 2021). Nevertheless, one can question whether such large gifts do not give the donor too much influence.

The Dutch government has recently adopted a law on party finance (Rijksoverheid, 2022). Individuals are no longer allowed to give more than 100,000 euro a year to one political party. As an electoral cycle takes four years, this culminates in 400,000 euro a year. Moreover, a benefactor seems to be able to give 200,000 to two parties annually. Such gifts can happen; for instance, D66 and PvdD received their gift from the same entrepreneur (Parool, 2021).

The data show that new parties were able to claim certain issues. Farmers' party BBB claimed agriculture, and spent more on this topic than the CDA, which is traditionally popular among farmers. Radical rightwing party JA21 claimed the issue of migration, largely because issue owner PVV does not buy many online advertisements. Finally, the new VOLT party was less successful in claiming issues. Volt was the fifth biggest spender on ads about climate. In line with Dobber et al. (2017), the data show how social media advertising can play an empowering role for smaller parties that struggle to get visibility through the traditional media. For example, smaller parties often struggle to get attention on TV in The Netherlands. Smaller parties are rarely invited for debates on national television. In some cases, smaller parties were ignored, and in other cases smaller parties had their own debates. However, new parties were not invited to these smaller party debates. Moreover, research has shown that news coverage favors the larger parties (Kostadinova, 2017), and is subject to horse race, conflict and campaign strategy news (Ergün and Karsten, 2019). New parties thus partly rely on social media advertising for visibility, however our data do not allow us to interpret the exact objectives for why parties advertise on these issues.

When we zoom in to the level of consideration set our data suggest that there are few contested issues: parties within such sets rarely aim to distinguish themselves from the other parties in the set. In the radical right consideration set, there are no contested issues. In the rightist consideration set, only climate, healthcare and transportation are contested. In the Christian consideration set, CDA owns each issue. In the leftist consideration set, finally, housing and healthcare are clearly contested. Economy, education and foreign affairs are contested to a lesser extent. Remarkably, where social democrat party PvdA claims no issue on the national stage, it is relatively unchallenged on issues of civil rights and social welfare in the leftist consideration set.

Research has shown that on the national level, Dutch citizens perceive housing and healthcare as two of the most important societal problems (only trumped by climate; Van der Meer & Damstra, 2022). Moreover, Dutch citizens rarely associate housing, and to a lesser degree healthcare, with any party (Van der Meer & Damstra, 2022). Indeed, our study shows that housing is one of the most prominent issues in terms of spending, but no party was able to claim that issue on the national level.

Lastly, we give some suggestions for further research. This study is limited in scope, as it focuses only on advertisements about one or multiple issues. Political parties also bought ads on Meta that were not about an issue. Thus, the results of this study cannot be used to make inferences about the general campaign. Moreover, as this study focuses on Meta, it does not discuss potential issue ads on other social media platforms. Future research might want to include other online platforms that enable political advertising. Moreover, online advertising does not occur in a vacuum: campaigns also advertise via traditional channels, and news media report on political parties and their issues throughout the campaign.

In addition, future studies could research issue ads while also researching the targeting strategies of the parties. As Meta (and to a lesser degree Google) provides political advertising with far-reaching targeting options, combining information about the issues advertised with the people targeted (i.e. gender, age brackets and geographical region) could present insights not only into what parties are communicating about, but also to whom. Future research could also investigate the issues of political ads further by looking at how political parties frame issues in their ads, in relation to their issue ownership. Such insights help us understand the online campaign, and monitor and flag potential undue influence, such as the influence of rich benefactors on the Dutch national election campaign. Finally, since the findings of this study suggest that political campaigns have different strategies for online advertising than traditional advertising, future research could compare traditional advertising with online advertising through the lens of issue ownership theory.

Funding and conflicts of interests

The authors have no funding to declare. The authors declare to have no conflict of interest.

References

Abbe, O.G., Goodliffe, J., Herrnson, P.S., Patterson and K.D. (2003) 'Agenda Setting in Congressional Elections: The Impact of Issues and Campaigns on Voting Behavior', Political Research Quarterly, 56(4), pp. 419–430. Available at: https://doi.org/10.1177/106591290305600404.

Adams, J., Ezrow, L. and Somer-Topcu, Z. (2011) 'Is Anybody Listening? Evidence That Voters Do Not Respond to European Parties' Policy Statements During Elections', American Journal of Political Science, 55(2), pp. 370–382. Available at: https://doi.org/10.1111/j.1540-5907.2010.00489.x.

Adams, J., Ezrow, L. and Somer-Topcu, Z. (2014) 'Do voters respond to party manifestos or to a wider information environment? An analysis of mass-elite linkages on European integration', American Journal of Political Science, 58(4), pp. 967–978. Available at: https://doi.org/10.1111/ajps.12115.

Andersen, R., Tilley, J. and Heath, A.F. (2005) 'Political knowledge and enlightened preferences: Party choice through the electoral cycle', British Journal of Political Science, 35(2), pp. 285–302. Available at: https://doi.org/10.1017/S0007123405000153.

Andeweg, R.B., De Winter And, L. and Müller, W.C. (2008) 'Parliamentary Opposition in Post-Consociational Democracies: Austria, Belgium and the Netherlands', The Journal of Legislative Studies, 14(1–2), pp. 77–112. Available at: https://doi.org/10.1080/13572330801921034.

Anstead, N. (2017) 'Data-Driven Campaigning in the 2015 United Kingdom General Election', International Journal of Press/Politics, 22(3), pp. 294–313. Available at: https://doi.org/10.1177/1940161217706163.

Banda, K.K. (2016) ‘Issue Ownership, Issue Positions, and Candidate Assessment’, Political Communication, 33(4), pp. 651–666. Available at: https://doi.org/10.1080/10584609.2016.1192569.

Béchara, H., Herzog, A., Jankin, S. and John, P. (2021) ‘Transfer learning for topic labeling: Analysis of the UK House of Commons speeches 1935–2014’, Research & Politics, 8(2), p. 20531680211022210. Available at: https://doi.org/10.1177/20531680211022206.

Bischof, D. and Senninger, R. (2018) ‘Simple politics for the people? Complexity in campaign messages and political knowledge’, European Journal of Political Research, 57(2), pp. 473–495. Available at: https://doi.org/10.1111/1475-6765.12235.

Bos, L., Lefevere, J. M., Thijssen, R., & Sheets, P. (2017). The impact of mediated party issue strategies on electoral support. Party Politics, 23(6), 760–771. Available at: doi.org/10.1177/1354068815626603

Chou, H.-Y. and Lien, N.-H. (2011) ‘What does a negative political ad really say? The effects of different content dimensions’, Journal of Marketing Communications, 17(4), pp. 281–295. Available at: https://doi.org/10.1080/13527260903546213.

Comparative Agendas Project (n.d.). https://www.comparativeagendas.net/netherlands

Coppock, A., Hill, S.J. and Vavreck, L. (2020) ‘The small effects of political advertising are small regardless of context, message, sender, or receiver : Evidence from 59 real-time randomized experiments’, pp. 1–7.

Dobber, T., Trilling, D. Helberger, N. and de Vreese, C. (2017) ‘Two crates of beer and 40 pizzas: the adoption of innovative political behavioural targeting techniques’, Internet Policy Review, 6(4), pp. 1–25. Available at: https://doi.org/10.14763/2017.4.777.

Dobber, T., Trilling, D. Helberger, N. and de Vreese, C. (2018) ‘Spiraling downward: The reciprocal relation between attitude toward political behavioral targeting and privacy concerns’, New Media & Society, 21, pp. 1–20. Available at: https://doi.org/10.1177/1461444818813372.

Dobber, T. and Vreese, C. de (2022) ‘Beyond manifestos: Exploring how political campaigns use online advertisements to communicate policy information and pledges’, Big Data & Society, 9(1), p. 20539517221095430. Available at: https://doi.org/10.1177/20539517221095433.

Downs, A. (1957) An Economic Theory of Democracy.

Endres, K. and Panagopoulos, C. (2019) ‘Cross-pressure and voting behavior: Evidence from randomized experiments’, Journal of Politics, 81(3), pp. 1090–1095. Available at: https://doi.org/10.1086/703210.

Ergün, E. and Karsten, N. (2019) ‘Media logic in the coverage of election promises: comparative evidence from the Netherlands and the US’, Acta Politica [Preprint], (0123456789). Available at: https://doi.org/10.1057/s41269-019-00141-8.

Fowler, E.F., Franz, M.M, Martin, G.J, Peskowitz, Z. and Ridout, T.N. (2020) ‘Political Advertising Online and Offline’, American Political Science Review, pp. 1–20. Available at: https://doi.org/10.1017/S0003055420000696.

Geers, S. and Bos, L. (2017) ‘Priming Issues, Party Visibility, and Party Evaluations: The Impact on Vote Switching’, Political Communication, 34(3), pp. 344–366. Available at: https://doi.org/10.1080/10584609.2016.1201179.

Geys, B. (2012) ‘Success and failure in electoral competition: Selective issue emphasis under incomplete issue ownership’, Electoral Studies, 31(2), pp. 406–412. Available at: https://doi.org/10.1016/j.electstud.2012.01.005.

Green-Pedersen, C. (2007). The Growing Importance of Issue Competition: The Changing Nature of Party Competition in Western Europe. Political Studies, 55(3), 607–628. Available at: https://doi.org/10.1111/j.1467-9248.2007.00686.x

Haenschen, K. (2022) ‘The Conditional Effects of Microtargeted Facebook Advertisements on Voter Turnout’, Political Behavior, pp. 409–433. Available at: https://doi.org/10.1215/03616878-8893529.

Haenschen, K. and Jennings, J. (2019) ‘Mobilizing Millennial Voters with Targeted Internet Advertisements: A Field Experiment’, Political Communication, 36, pp. 357–375. Available at: https://doi.org/10.1080/10584609.2018.1548530.

Hillygus, D.S. and Shields, T.G. (2009) The persuadable voter: Wedge issues in presidential campaigns. Princeton: Princeton University Press.

Jamieson, K.H. (2013) ‘Messages, micro-targeting, and new media technologies’, Forum (Germany), 11(3), pp. 429–435. Available at: https://doi.org/10.1515/for-2013-0052.

Kanne, P. & Van de Koppel, M. (2023, 16 September). I&O-zetelpeiling: NSC, PvdA/GL en VVD samen aan kop. I&O Research. Retrieved from https://www.ioresearch.nl/actueel/verkiezingsstrijd-ligt-open-nsc-pvda-gl-en-vvd-samen-aan-kop/

Kiesraad (2021). Officiële uitslag Tweede Kamerverkiezing 17 maart 2021. 26 March, https://www.kiesraad.nl/actueel/nieuws/2021/03/26/officiele-uitslag-tweede-kamerverkiezing-17-maart-2021

Kostadinova, P. (2017) ‘Party pledges in the news: Which election promises do the media report?’, Party Politics, 23(6), pp. 636–645. Available at: https://doi.org/10.1177/1354068815611649.

Kreiss, D. (2016) Prototype Politics: Technology-intensive campaigning and the data of democracy.

Krippendorff, K.H. (2004) Content Analysis: An Introduction to Its Methodology.

Kruikemeier, S., Vermeer, S. Metoui, N., Dobber, T. and Zarouali, B. (2022) ‘(Tar)getting you: The use of online political targeted messages on Facebook’, Big Data & Society, 9(2), p. 20539517221089624. Available at: https://doi.org/10.1177/20539517221089626.

Kruschinski, S. and Haller, A. (2017) ‘Restrictions on data-driven political microtargeting in Germany’, Internet Policy Review, 6(4), pp. 1–23. Available at: https://doi.org/10.14763/2017.4.780.

Lavigne, M. (2020) ‘Strengthening ties: The influence of microtargeting on partisan attitudes and the vote’, Party Politics, (August 2019), p. 135406882091838. Available at: https://doi.org/10.1177/1354068820918387.

Leerssen, P., Ausloos, J., Zarouali, B., Helberger, N. and de Vreese, C. (2019) ‘Platform ad archives: Promises and pitfalls’, Internet Policy Review, 8(4), pp. 1–21. Available at: https://doi.org/10.14763/2019.4.1421.

Leerssen, P., Dobber, T., Helberger, N. and de Vreese, C. (2021) ‘News from the ad archive: how journalists use the facebook ad library to hold online advertising accountable’, Information, Communication & Society, 0(0), pp. 1–20. Available at: https://doi.org/10.1080/1369118x.2021.2009002.

McCombs, M.E. and Shaw, D.L. (1972) ‘The Agenda-Setting Function of Mass Media’, The Public Opinion Quarterly, 36(2), pp. 176–187.

van der Meer, T. and Damstra, A. (2022) ‘Associative issue ownership in a highly fragmented multiparty context: The Netherlands (2021)’, Acta Politica [Preprint]. Available at: https://doi.org/10.1057/s41269-022-00274-3.

NOS (2021). Donatie van 1,2 miljoen aan CDA door ondernemer Van der Wind roept vragen op. https://nos.nl/artikel/2385144-donatie-van-1-2-miljoen-aan-cda-door-ondernemer-van-der-wind-roept-vragen-op

NPO Radio (2021). Politieke partijen afhankelijker van donaties in 2021. https://www.nporadio1.nl/nieuws/politiek/25f3609a-d18b-4ea6-8166-8d4e2da00a5d/politieke-partijen-afhankelijker-van-donaties-in-2021

Parool (2021). Techondernemer schenkt D66 miljoen euro, PvdD krijgt meer dan 3 ton. https://www.parool.nl/nederland/techondernemer-schenkt-d66-miljoen-euro-pvdd-krijgt-meer-dan-3-ton~b9044379/

Passonneau, R., Habash, N.Y. and Rambow, O.C. (2006) ‘Inter-annotator agreement on a multilingual semantic annotation task’. Available at: https://doi.org/10.7916/D8FB5BDX.

Petrocik, J.R. (1996) ‘Issue Ownership in Presidential Elections , with a 1980 Case Study’, American Journal of Political Science, 40(3), pp. 825–850.

Petrocik, J.R., Benoit, W.L. and Hansen, G.J. (2003) ‘Issue Ownership and Presidential Campaigning, 1952-2000’, Political Science Quarterly, 118(4), pp. 599–626.

Rekker, R. and Rosema, M. (2019) ‘How (often) do voters change their consideration sets?’, Electoral Studies, 57(October 2017), pp. 284–293. Available at: https://doi.org/10.1016/j.electstud.2018.08.006.

Rijksoverheid (2022). Wet financiering politieke partijen. https://wetten.overheid.nl/BWBR0033004/2023-01-01

Turow, J., Carpini, M.X.D., Draper, N. and Howard-Williams, R., (2012) Americans Roundly Reject Tailored Political Advertising, Annenberg School for Communication, University of Pennsylvania.

Votta, F., Noroozian, A., Dobber, T., Helberger, N. and de Vreese, C. (2023) ‘Going Micro to Go Negative?: Targeting Toxicity using Facebook and Instagram Ads’, Computational Communication Research, 5(1), pp. 1–50. Available at: https://doi.org/10.5117/CCR2023.1.001.VOTT.

Walgrave, S. and De Swert, K. (2004) ‘The Making of the (Issues of the) Vlaams Blok’, Political Communication, 21(4), pp. 479–500. Available at: https://doi.org/10.1080/10584600490522743.

Walgrave, S., Lefevere, J. and Nuytemans, M. (2009) ‘Issue Ownership Stability and Change: How Political Parties Claim and Maintain Issues Through Media Appearances’, Political Communication, 26(2), pp. 153–172. Available at: https://doi.org/10.1080/10584600902850718.

Walgrave, S., Tresch, A. and Lefevere, J. (2015) ‘The Conceptualisation and Measurement of Issue Ownership’, West European Politics, 38(4), pp. 778–796. Available at: https://doi.org/10.1080/01402382.2015.1039381.

Zarouali, B., Dobber, T., De Pauw, G. and de Vreese, C. (2020) ‘Using a Personality-Profiling Algorithm to Investigate Political Microtargeting: Assessing the Persuasion Effects of Personality-Tailored Ads on Social Media’, Communication Research, pp. 1–26. Available at: https://doi.org/10.1177/0093650220961965.

Zuiderveen Borgesius, F.J., Möller, J., Kruikemeier, S., Fathaigh, R.Ó., Irion, K., Dobber, T., Bodo, B. and de Vreese, C (2018) ‘Online political microtargeting: Promises and threats for democracy’, Utrecht Law Review, 14(1), pp. 82–96. Available at: https://doi.org/10.18352/ulr.420.

Appendix A. Party Information

Table 3. List of Dutch Political Parties that gained at least one seat in the 2021 Dutch general election

Party	**Ideology**	**Political leader**[5]	**Website**
50+	Pensioners' interests	Liane den Haan	https://50pluspartij.nl/
BBB	Agrarianism	Caroline van der Plas	https://boerburgerbeweging.nl/
BIJ1	Intersectionality	Sylvana Simons	https://bij1.org/
CDA	Christian democracy	Wopke Hoekstra	https://www.cda.nl/
ChristenUnie	Christian democracy	Gert-Jan Segers	https://www.christenunie.nl/
D66	Progressive liberalism	Sigrid Kaag	https://d66.nl/
DENK	Multiculturalism	Farid Azarkan	https://www.bewegingdenk.nl/
FvD	National conservatism	Thierry Baudet	https://fvd.nl/
GroenLinks	Green politics	Jesse Klaver	https://www.groenlinks.nl/
JA21	Conservative liberalism	Joost Eerdmans	https://ja21.nl/
PvdA	Social democracy	Lilianne Ploumen	https://www.pvda.nl/
PvdD	Animal rights	Esther Ouwehand	https://www.partijvoordedieren.nl/
PVV	Right-wing populism	Geert Wilders	https://pvv.nl/
SGP	Conservative Calvinism	Kees van der Staaij	https://sgp.nl/
SP	Socialism	Lilian Marijnissen	https://www.sp.nl/
Volt	Progressive liberalism	Laurens Dassen	https://voltnederland.org/
VVD	Conservative liberalism	Mark Rutte	https://www.vvd.nl/

[5] During the 2021 Dutch general election.

https://doi.org/10.33621/jdsr.v7i154885

Table 4. Issues and impressions per party

	50+	BBB	BIJ1	CDA	CU	D66
Agriculture	0	4024982	0	886472	0	9500
Civil Rights	0	172498	753984	9749245	400992	870982
Climate	0	1927494	112498	716935	247994	4599916
Defense	0	0	0	187498	500	0
Economy	188996	1109998	0	7999384	44998	57498
Education & Culture	0	65000	251498	2410852	463494	3707892
Foreign Affairs	17500	0	0	0	6498	0
Government	0	89999	5500	3678758	61997	35998
Healthcare	514991	42500	82997	1098908	275494	2346452
Housing	920484	3345482	57997	6318088	34996	1491480
Law & Order	17500	75000	44999	1558848	91499	43998
Migration	0	137500	0	835420	5999	0
Social Welfare	1627482	264998	25999	10652300	182994	129494
Transportation	0	314998	0	161494	11494	197496
Total	3286953	11570449	1335472	46254202	1828949	13490706

	DENK	FvD	GL	JA21	PVV	PvdA
Agriculture	0	678497	117999	0	0	8498
Civil Rights	494484	1913494	1098982	0	0	7699250
Climate	0	691494	3965967	139999	2500	210486
Defense	0	1216996	226499	0	0	0
Economy	147991	2027496	1318494	2117994	0	3961372
Education & Culture	184984	438998	582492	30999	0	2121796
Foreign Affairs	48494	0	42500	0	0	145993
Government	0	384998	537996	0	0	275990
Healthcare	111994	5412490	2163490	2117994	0	3565773
Housing	128991	889998	1090988	0	0	3889239
Law & Order	0	244998	500	2123496	0	17496
Migration	45997	769997	9498	4361488	0	8500
Social Welfare	16997	1539996	1029491	0	0	5341562
Transportation	51494	342498	914994	0	0	76993
Total	1231426	16551950	13099890	10891970	2500	27322948

	PvdD	**SGP**	**SP**	**VOLT**	**VVD**
Agriculture	2353484	68996	0	0	49498
Civil Rights	878994	84992	5833476	159980	2336984
Climate	5534443	128996	162993	1564479	2088470
Defense	0	0	0	0	173991
Economy	427492	69998	777986	0	494962
Education & Culture	52498	59498	6498	9999	31994
Foreign Affairs	0	0	0	0	0
Government	3498	27500	1064496	0	115994
Healthcare	1364992	73995	1919991	0	1867464
Housing	686984	66997	3597976	2500	2538408
Law & Order	0	17500	446497	0	618980
Migration	0	8500	0	0	1500
Social Welfare	17500	4999	1267488	0	227494
Transportation	256992	32500	242499	0	70990
Total	11576877	644471	15319900	1736958	10616729

https://doi.org/10.33621/jdsr.v7i154885

Appendix B. Money pent per issue

Table 5. Money spent per issue per party (in EUR)

	50+	BBB	BIJ1	CDA	CU	D66	DENK	FvD	GL
Agriculture	0,00	16.382,00	0,00	7.322,00	0,00	50,00	0,00	2.497,00	299,00
Civil Rights	0,00	848,00	5.634,00	119.745,00	4.592,00	3.232,00	2.734,00	9.044,00	7.932,00
Climate	0,00	8.744,00	148,00	8.335,00	2.744,00	16.066,00	0,00	1.894,00	24.467,00
Defense	0,00	0,00	0,00	448,00	50,00	0,00	0,00	2.546,00	799,00
Economy	846,00	5.798,00	0,00	89.434,00	898,00	448,00	891,00	9.696,00	8.894,00
Education & Culture	0,00	250,00	948,00	20.702,00	4.594,00	19.042,00	1.584,00	1.248,00	3.742,00
Foreign Affairs	150,00	0,00	0,00	0,00	148,00	0,00	544,00	0,00	250,00
Government	0,00	599,00	50,00	43.258,00	497,00	198,00	0,00	1.048,00	2.846,00
Healthcare	1.791,00	250,00	797,00	10.458,00	3.044,00	12.452,00	594,00	21.440,00	10.290,00
Housing	3.234,00	14.382,00	597,00	64.238,00	596,00	16.480,00	1.091,00	2.448,00	6.838,00
Law & Order	150,00	550,00	399,00	18.098,00	499,00	398,00	0,00	448,00	50,00
Migration	0,00	550,00	0,00	9.770,00	99,00	0,00	297,00	2.097,00	148,00
Social Welfare	5.932,00	848,00	99,00	112.700,00	1.644,00	1.044,00	297,00	6.096,00	5.491,00
Transportation	0,00	1.348,00	0,00	1.844,00	644,00	1.346,00	644,00	948,00	7.344,00

	JA21	PVV	PvdA	PvdD	SGP	SP	VOLT	VVD
Agriculture	0,00	0,00	148,00	15.434,00	396,00	0,00	0,00	248,00
Civil Rights	0,00	0,00	85.442,00	7.194,00	1.192,00	41.275,00	2.130,00	5.384,00
Climate	299,00	42,00	1.636,00	54.943,00	846,00	1.393,00	9.179,00	9.520,00
Defense	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1.091,00
Economy	6.344,00	0,00	16.422,00	3.842,00	248,00	5.236,00	0,00	4.312,00
Education & Culture	199,00	0,00	24.546,00	548,00	448,00	248,00	99,00	594,00
Foreign Affairs	0,00	0,00	693,00	0,00	0,00	0,00	0,00	0,00
Government	0,00	0,00	1.440,00	148,00	250,00	5.395,00	0,00	1.244,00
Healthcare	6.344,00	0,00	27.473,00	11.792,00	695,00	8.691,00	0,00	7.664,00
Housing	0,00	0,00	31.339,00	4.634,00	597,00	23.126,00	50,00	21.958,00
Law & Order	6.846,00	0,00	346,00	0,00	50,00	1.297,00	0,00	2.430,00
Migration	14.038,00	0,00	50,00	0,00	50,00	0,00	0,00	50,00
Social Welfare	0,00	0,00	53.812,00	150,00	99,00	10.788,00	0,00	1.044,00
Transportation	0,00	0,00	893,00	1.742,00	250,00	799,00	0,00	990,00

Appendix C. Issue word lists

Table 6. Issue word lists

Agriculture	Civil Rights	
platteland	eerlijk	genderidentiteit
boer	samenleving	grondrecht
landbouw	gelijk	journalist
bos	waarde	religieus
voedsel	norm	islam
boerenbedrijf	vrouw	integratie
veehouderij	burger	meningsuiting
agrarisch	vrijheid	mensenrechten
koe	fatsoen	geestelijk
wei	recht	queer
bestrijding	ongelijkheid	seks
boerennatuur	rechtvaardig	nationaliteit
visser	openbaar	trans
noaberschap	kansengelijkheid	etnisch
product	leeftijd	minderheid
weer	intimidatie	stemrecht
voeding	discriminatie	non-binair
landbouwgrond	traditie	antisemitisme
visserij	behandeling	transgender
vlees	racisme	zelfbeschikkingsrecht
dierenwelzijn	afkomst	kiesrecht
halal	religie	regenboogvlag
kip	toeslagenaffaire	gehandicapt
bestemmingsplan	gelijkwaardigheid	antiracisme
transport	seksueel	menswaardig
intensief	geaardheid	godsdienst
verkoop	emancipatie	seksualiteit
milieuvriendelijk	diversiteit	extremisme
oogst	cultureel	vrouwenemancipatie
stal	islamitisch	oeigoeren
voedselproductie	commercieel	gendergerelateerd
appel	rechtvaardigheid	lhbtqia+-gemeenschap
boerin	taal	lhbtqia+-kwestie
melkveehouderij	medisch	blind
veestapel	vrouwendag	openbaarheid
stankoverlast		privacy
varken		
biologisch		
import		
ingredient		
keurmerk		
plattelandsbeleid		
plattelandspartij		
voedselproducent		
boerenfamilie		
boerenkeukentafel		
boerenpartij		
koeienboer		
landbouwbeleid		
slachterij		
vee		

https://doi.org/10.33621/jdsr.v7i154885

Climate			Defense
duurzaam	luchtkwaliteit	milieuprobleem	defensie
groen	plastic	gas	vrede
klimaat	vervuilende		dreiging
wereld	transport		landmacht
natuur	klimaatdoelstelling		oud-commandant
klimaatverandering	klimaatbedrog		oorlog
energie	stikstofwet		aanval
schoon	otter		missie
gebied	stroom		militair
kernenergie	klimaatvriendelijk		kernwapen
omgeving	windmolenpark		leger
planeet	klimaataanpak		conflict
duurzaamheid	droogte		materieel
klimaatakkoord	brandstof		kazerne
windmolen	zonne-energie		bondgenoot
milieu	klimaatproblem		krijgsmacht
klimaatcrisi	stikstofuitstoot		vn
leefomgeving	vergroening		veteraan
grond	chemisch		soldaat
boom	klimaatplan		afghanistan
bodem	energieopwekking		uitrusting
klimaatbeleid	kust		wereldoorlog
lucht	elektriciteit		libie
zee	zoutkoepel		Wapen
landschap	fijnstof		herbewapening
biomassa	natuurbeheer		ontwapening
energietransitie	kernafval		wapenwedloop
water	nucleair		ontwapening
park	stikstofdebat		kernbom
klimaatactie	beschermen		agressor
afval	olie		joegoslavie
verkeer	stikstofregel		irak
bescherming	alcohol		defensie-industrie
zonnepaneel	energievoorziening		officierenvereniging
mega-zonnepark	gaswinning		wapenembargo
ruimtelijk	radioactief		vn-raad
natuurgebied	energiebesparing		vliegbasis
beek	groenstrook		alliantie
zonnepark	milieubeleid		corps
plant	milieuvervuiling		infiltratie
energiemix	onbewoonbaar		jihadistenstrijder
controle	overstroming		navo-bondgenoot
windpark	stikstofdiscussie		veiligheidssector
klimaatalarm	duurzaamheidsagenda		veteranendag
energiebron	uranium		somalie
klimaatwet	luchtvervuiling		knil-militair
terrein	bestrijdingsmiddel		defensieplek
co2	broeikasgas		
vervuiling	carpool		
waterschap	energiebehoefte		
klimaatverkiezing	grondwater		
stikstof	habitat		
klimaatoplossing	leefklimaat		
uitstoot	mega0zonnepark		
rivm	milieubeweging		
klimaatneutraal	ontbossing		
experiment	recycling		
diersoort	regen		
stof	watervoorziening		
klimaatplann	co2-probleem		
windenergie	klimaatdoemdenker		
bestemmingsplan	energiebeleid		

Economy		Education & Culture	
betaalbaar	bank	onderwijs	museum
economie	industrie	leenstelsel	erfgoed
ondernemer	werkloosheidsuitkering	school	radio
salaris	verlies	basisbeurs	vmbo
belasting	sluiting	cultuur	kunstenaar
baan	bedrijventerrein	student	basisonderwijs
geld	ozb	openbaar	cursus
minimumloon	zzp'er	studieschuld	techniek
loon	handel	technologie	schooljaar
bedrijf	investeerder	onderzoek	onderwijshuisvesting
armoede	winkelier	docent	televisie
investering	eigendom	leerling	communicatie
ontwikkeling	bedrijfsleven	media	scholier
economisch	zzp	kerk	immigrant
middeninkomen	werkloosheid	innovatie	ek
kost	winkelstraat	stage	sportvereniging
arbeidsmarkt	klant	internet	collegegeld
schuld	ondernemersklimaat	leerkracht	pers
herstel	kleinbedrijf	sport	middelbaar
winkel	commerciel	studie	lerarentekort
goedkoop	retail	ondersteuning	mobiel
werknemer	aandeelhouder	god	onderwijsplann
begroting	aandeel	krant	cultuursector
speculant	onderneming	religie	beroepsonderwijs
financiel	toezicht	geloof	eindexamen
portemonnee	ondernemend	moslim	schoolbestuur
markt	verkoop	basisschool	universitair
groeifonds	vastgoed	kerst	klaslokaal
schuldenvrij	dividend	universiteit	examenleerling
mkb	btw	christen	joods
financieel	faillissement	bibliotheek	scholing
belegger	onderneemster	klas	godsdienst
hypotheek	consument	mbo	cultuurbeleid
belastinggeld	fraude	opleiding	technologisch
werkgarantie	ondernemersspreekuur	cultureel	digitalisering
groei	accijns	moslimhaat	glasvezel
belastingverlaging	bv	islamitisch	expertise
last	btw-verhoging	kunst	vwo
staatsschuld	klantenkring	taal	cbs
bezuiniging	koopkracht	theater	ict
winst	verzekeraar	oranje	wetenschappelijk
werkgelegenheid	ondernemersprij	wetenschapper	duo
marktwerking	krediet	christelijk	allochtoon
budget	vestigingsklimaat	katholiek	cijferlijst
werkgever	fiscaal	voetbal	collecte
krimpgebied	investeringsruimte	hbo	computer
eigenaar	overheidssteun	religieus	cultuuraanbod
toerisme	overname	instelling	cultuurhistorisch
ondernemerschap	groeiregio	islam	havo
		kerstdag	kersttoespraak
		kerstboodschap	loting
		asielzoeker	ruimtevaart
		les	sportbeleid

Foreign Affairs	Government
europees	overheid
vluchteling	burger
internationaal	minister
trump	rechtsstaat
terrorisme	regering
amerikaans	provincie
donald	rijk
eu	referendum
ontwikkelingsland	grondwet
diplomaat	burgemeester
zuid-europa	bestuur
turkije	parlement
midden-oosten	bestuurlijk
rusland	transparantie
azerbeidzjan	publiek
joe	stadsbestuur
biden	bezuiniging
buitenlands	bestuurder
brussel	feestdag
oorlog	bureaucratie
armenie	stemadvies
onderdrukking	verkiezingstijd
armeens	grondrecht
europarlementarier	minister-president
humanitair	brandweer
us	staatssecretaris
israel	koning
handel	ambtenaar
ontwikkelingssamenwerking	kandidaatstelling
mensenrechten	koninklijk
conflict	ramp
duitsland	koningsdag
griekenland	koninkrijk
buitenland	grondwettelijk
oost-europa	overheidsdienst
kruis	commissaris
ambassadeur	kamerkandidat
afrika	gemeentewet
turks	begrafenis
diplomatie	bestuurslaag
erdogan	koningin
sanctie	decentralisatie
chinees	gemeentefonds
syrie	rampenbestrijding
japan	staatkundig
kanaalzone	staking
presidentsverkiezing	
koninkrijksrelaties	
import	
eu-lidmaatschap	
groot-brittannie	
dictatuur	
handelsakkoord	
israelisch	
ontwikkelingshulp	
ambassade	

https://doi.org/10.33621/jdsr.v7i154885

Healthcare		Housing	
crisis	experiment	woning	huurcrisi
zorg	arbeidsvoorwaarde	huis	wooncorporatie
kind	covid	betaalbaar	huurstijging
coronacrisi	coronaregel	straat	dakloos
corona	drug	regio	woonplicht
voorziening	medicijn	volkshuisvesting	kantoor
zorgpremie	vaccinatiestrategie	stad	ozb
gezond	lichamelijk	inwoner	huurbevriezing
lockdown	familielid	belasting	leegstand
kwetsbaar	verpleeghuis	wijk	woningbouwplan
abortus	langdurig	ruimte	woonlast
gezondheidszorg	zorgkosten	bouw	woningcrisi
mantelzorger	corona-herstelfonds	starter	bestemmingsplan
bijdrage	geboorte	dorp	verzorgingstehuis
behandeling	gehandicapt	platteland	dakloosheid
sport	coronafonds	gebied	verhuizing
gezondheid	vaccinatiepaspoort	woningmarkt	flat
ziekenhuis	vaccinatieplicht	huisjesmelker	huurhuis
volksgezondheid	specialist	wooncrisi	kamerverhuur
seksueel	overgewicht	omgeving	brandveilig
coronaviru	avondklokrell	bewoner	elektriciteit
vaccinatie	psychiatrisch	woningnood	vestiging
zorgverlener	who	huur	woningprijz
versoepeling	alcohol	prijs	verpaupering
patient	zorgverzekeraar	huurwoning	huurteam
preventie	aandoening	woningbouw	appartementengebouw
besmetting	ambulancepost	huurder	onleefbaarheid
ambulance	zorgtoeslag	leefomgeving	registratie
vaccin	ggd	randstad	stedeling
medisch	ic-capaciteit	leefbaarheid	verhuurdersvergunning
huisarts	operatie	gemeenschap	huurcontract
zorgmedewerker	verslaving	woningtekort	huurdersraadpleging
welzijn	verwijzing	hypotheek	binnenstedelijk
ziekte	verzekeraar	koopwoning	verhuurgedrag
wachtlijst	zorginstelling	nieuwbouwwoning	bejaardenhuis
persconferentie	intensive	premie	huisvestingsbeleid
personeel	thuiszorg	huisvesting	huurbescherming
psychisch	tandarts	binnenstad	huurdersbescherming
test	ambulancestandplaats	huurverhoging	huurplafond
vaccinatieprogramma	ambulancezorg	gebouw	middenhuurwoning
voeding	bloeddonatie	huurprijs	premiewoning
donatie	dienstverlening	woonplaats	stadsvernieuwing
mondkap	reikwijdte	dakloze	studentenwoning
voorlichting	revalidatie	starterswoning	woningvoorraad
coronacrisis	schadevergoeding	stedelijk	
abortuskliniek	hygiene		
rivm	vergoeding		
drank	virus		

Law & Order		Migration
drugscriminaliteit	drug	migratie
rechtsstaat	juridisch	grens
geweld	justitie	asiel
criminaliteit	onveiligheid	arbeidsmigratie
rechtvaardig	verkrachting	immigratie
veiligheid	drugsproductie	asielbeleid
vernieling	gerechtshof	immigratiebeleid
intimidatie	misbruik	integratie
geweldsincident	rechtszaak	asielzoeker
grondwet	om	migratieachtergrond
drugsafval	legaal	marrakesh
liquidatie	huwelijk	immigrant
steekpartij	cel	inburgering
terrorisme	misleiding	vreemdelingenzaak
overlast	fraude	asielsysteem
boete	loverboy	asielzoekerscentrum
politie	radicalisering	allochtoon
jeugdzorg	rechterlijk	remigratie
straatintimidatie	awb	niet-westers
preventie	advocatuur	immigratiepact
speed	bodycam	polen
xtc	bevoegdheid	midden-oosten
kraker	rechtspraak	afrika
rechtvaardigheid	wapenstok	moria
dreiging	wijkagent	vluchtelingenkamp
prostitutie	draagmoederschap	vluchteling
crimineel	gevangenis	pardon
straf	vrijlating	instroom
grondrecht	bordeel	migratiecrisis
uitbuiting	burgerlijk	moria-deal
kraak	drugsgeld	migratiepact
coffeeshop	drugshandel	selectiecentrum
onveilig	jeugdgevangenis	asielbelofte
inbreker	opsporing	libie
illegaal	taakstraf	turkije
extreem	veelpleger	azc
diefstal	wapenbezit	opvangplek
woningoverlast	wetboek	
agent	strafblad	
bedreiging	tuig	
cameratoezicht	gevangenisstraf	
camera	gebiedsverbod	

Social Welfare		Transportation	
samenleving	ziekte	weg	parkeerbeleid
werk	opvang	bereikbaar	verkeersprobleem
salaris	multinational	openbaar	randweg
sociaal	voedselbank	vervoer	rotonde
contract	werkloosheidsuitkering	verbinding	verkeersplan
minimumloon	vakbond	snelweg	buslijn
loon	armoedebestrijding	zee	verkeersdeelnemer
bedrijf	arbeidsvoorwaarde	lijn	fietsenstalling
armoede	werkloosheid	auto	tram
pensioen	verzorgingstehuis	trein	oversteekplaats
middenklasse	thuiswerk	luchtvaart	waterstaat
voorziening	vrijwilligerswerk	verkeer	dienstregeling
arbeid	bonus	bereikbaarheid	luchthaven
hulp	onderneming	ov	truckchauffeur
middeninkomen	basisinkomen	binnenstad	jachthaven
inkomen	samenhang	haven	verkeershufter
handicap	gehandicapt	kilometer	a27
compensatie	loopbaan	omwonende	hogesnelheidstrein
arbeidsmarkt	werkloos	fiets	infrastructureel
schuld	bijbaan	station	metro
leeftijd	pensioenfonds	mobiliteit	vertraging
aow	vrijwilligersorganisatie	knelpunt	scooter
tekort	tolerant	verkeersveiligheid	parkeerproblematiek
flexcontract	zieken	tarief	sneltrein
arbeidsmigratie	jeugdloon	infrastructuur	verkeersdrukte
vaardigheid	nabestaande	snelheid	verkeersstroom
stage	omscholing	maximumsnelheid	voertuig
werknemer	vergrijzing	km	vrachtwagen
zelfstandig	werkplek	verbindingsweg	sanering
financiel	ouderenbeleid	rondweg	scheepvaart
pensioenstelsel	zomerkamp	vliegtuig	intercity
ww	allochtoon	bus	130km/u
ondersteuning	bedrijfstak	reiziger	a35
maatschappelijk	beschuldiging	vliegveld	apk
arbeidsongeschiktheid	bouwwerkzaamheid	brug	autoparkeerplaats
uitkering	cao	schip	busdienst
schuldenvrij	huishouding	fietser	busmaatschappij
loonkloof	jeugdwerkloosheid	tunnel	dienstverlening
kinderopvang	overwerk	voetganger	foutparkerend
bijstand	schuldsanering	parkeerplaats	n209
jeugd	seizoensarbeid	vervoersteun	parkeermogelijkheid
jeugdzorg	staking	automobilist	parkeerruimte
topinkomen	werkomstandigheid	verkeersslachtoffer	perron
fnv	ziektekosten	verbreding	sneltram
subsidie	zwangerschapsverlof	fusie	tolheffing
werkgelegenheid	sollicitatie	ov-idee	verkeerschaos
premie	werkvloer	transport	verkeersveiligheidsoogpunt
werkgever	weeskind	parkeergarage	verkeersveiligheidsproblem
welzijn		ontsluitingsweg	busvervoer
		ramp	parkeerbelasting
		ongeluk	spoortunnel
		parkeerplek	garage
		file	parkeergeld
		asfalt	

https://doi.org/10.33621/jdsr.v7i154885

Appendix D. Krippendorff's Alpha for each theme

In the method section, we discussed the validation of the model. We computed the multi-label Krippendorff's alpha to measure the inter-coder reliability. In this appendix, we provide an inter-coder reliability measurement for each individual issue.

For each issue, we computed the Krippendorff's alpha (using a nominal distance function).

As not all issues were significantly represented in the random sample of 300 ads used for validation, we only computed the Krippendorff's alpha for an issue if at least 15 ads (i.e. 5% of the sample) were considered to be about the issue by either coder (the total is provided in the "# ads" column).

Issue	# ads	Krippendorff's Alpha
Agriculture	10	-
Civil Rights	101	0.9
Climate	21	0.88
Defense	1	-
Economy	46	0.85
Education & Culture	20	0.84
Foreign Affairs	1	-
Government	17	0.82
Healthcare	35	0.83
Housing	54	0.76
Law & Order	12	-
Migration	2	-
Social Welfare	58	0.83
Transportation	7	-

https://doi.org/10.33621/jdsr.v7i154885

Appendix E. Spending per Issue per Party

Figure 7. Spending per Issue by CDA & 50Plus (in EUR)

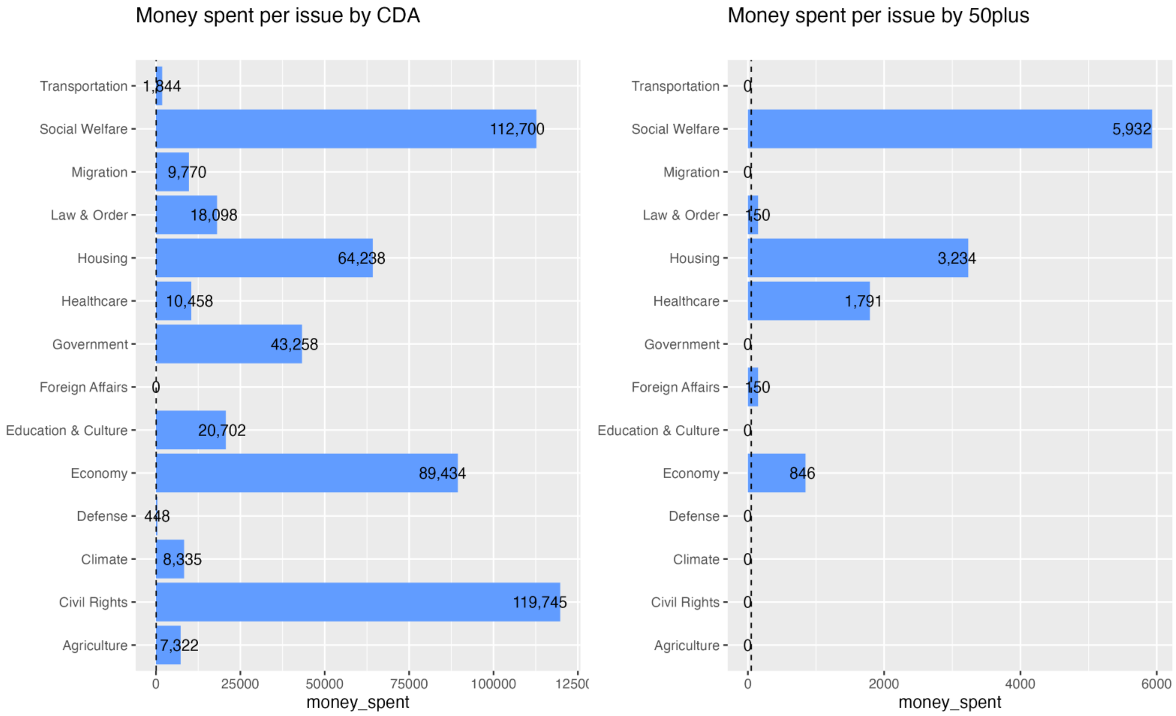


Figure 8. Spending per Issue by BBB & Bij1 (in EUR)

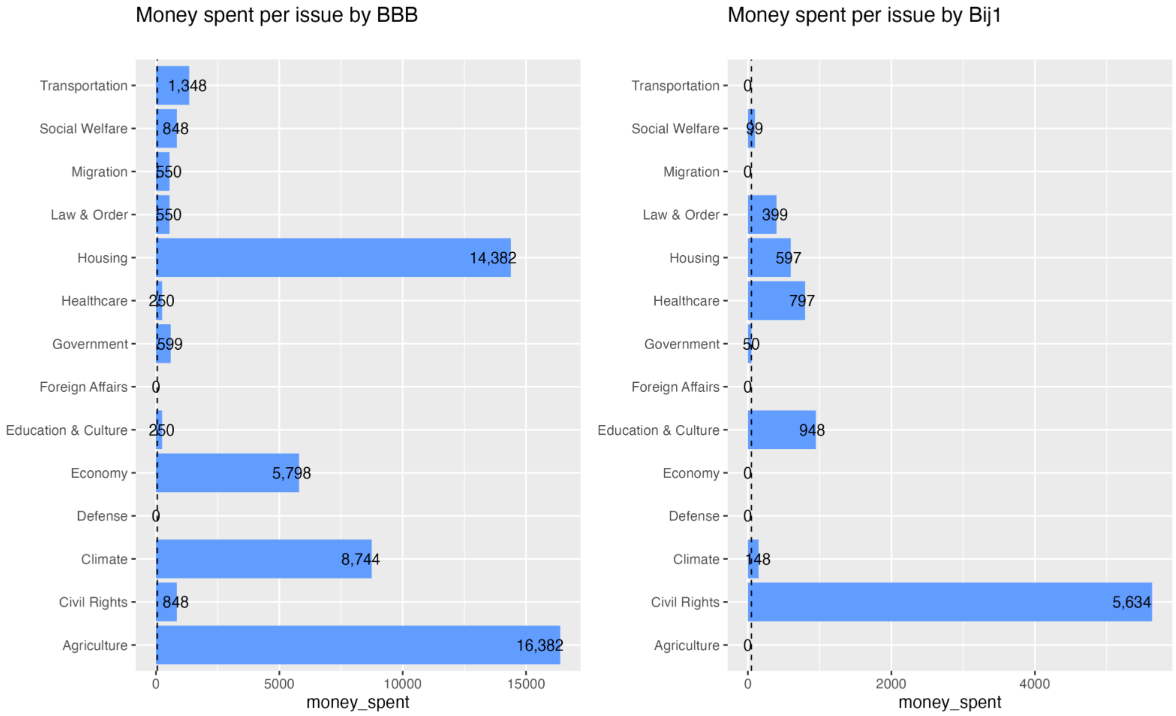
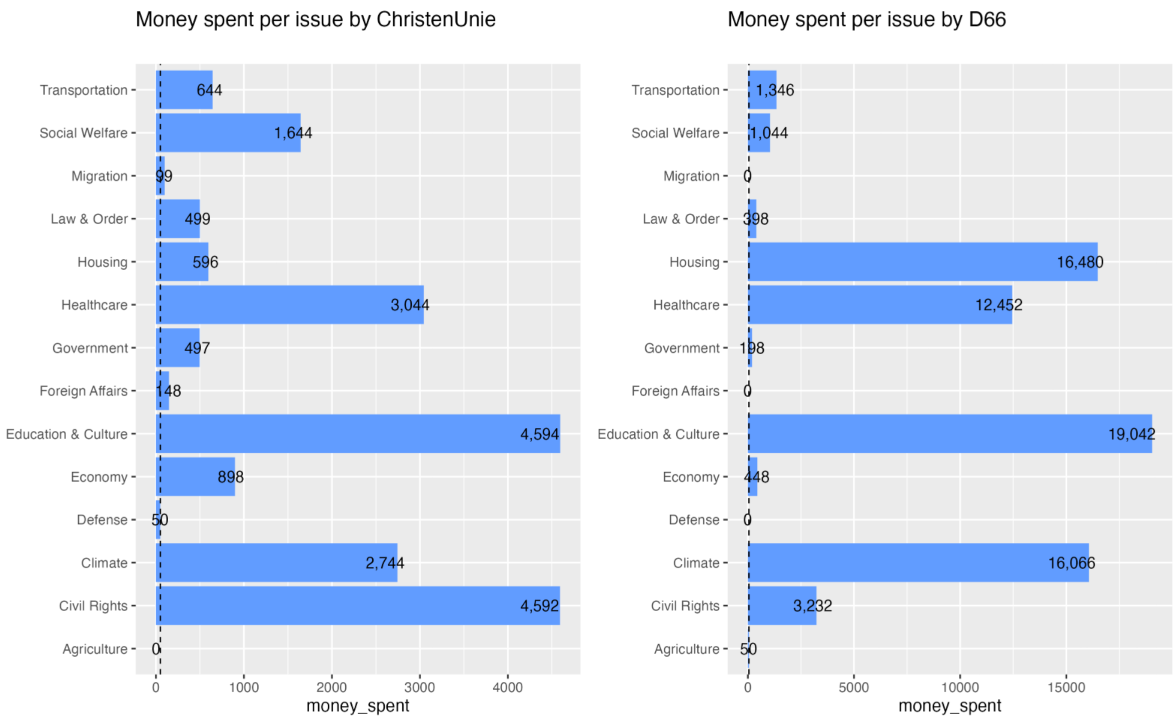


Figure 9. Spending per Issue by ChristenUnie & D66 (in EUR)

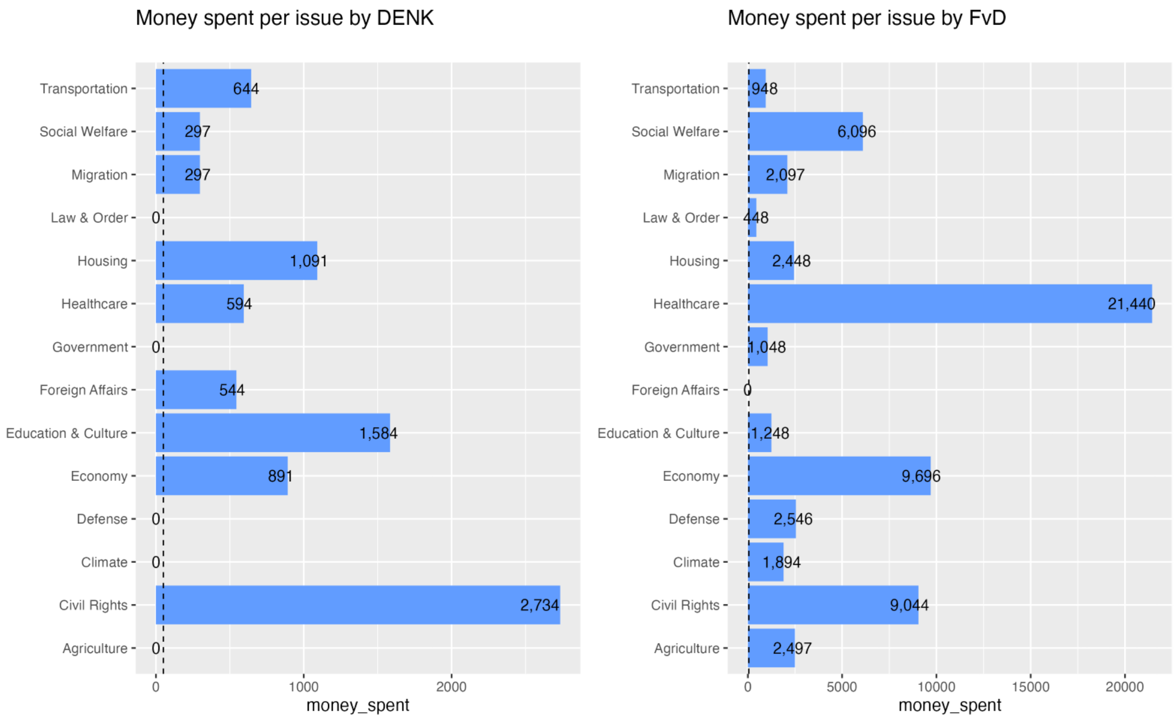
https://doi.org/10.33621/jdsr.v7i154885



Money spent per issue by ChristenUnie

Money spent per issue by D66

Figure 10. Spending per Issue by Denk & FvD (in EUR)



Money spent per issue by DENK

Money spent per issue by FvD

Figure 11. Spending per Issue by JA21 & PvdA (in EUR)

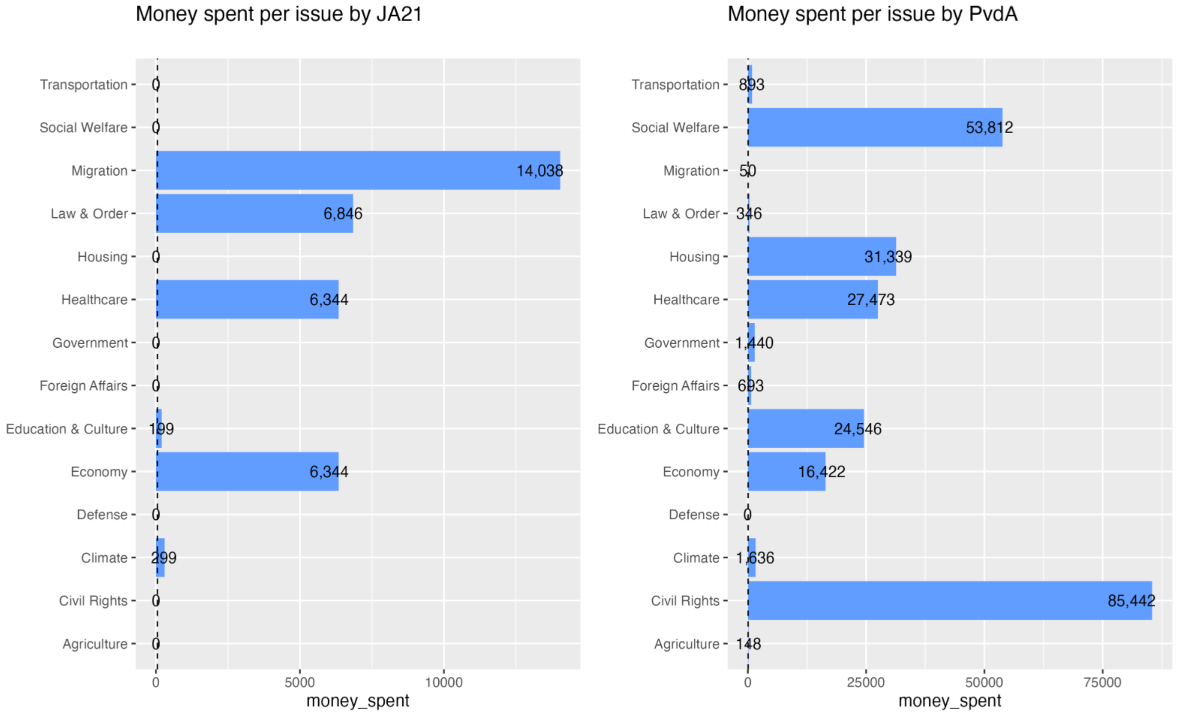
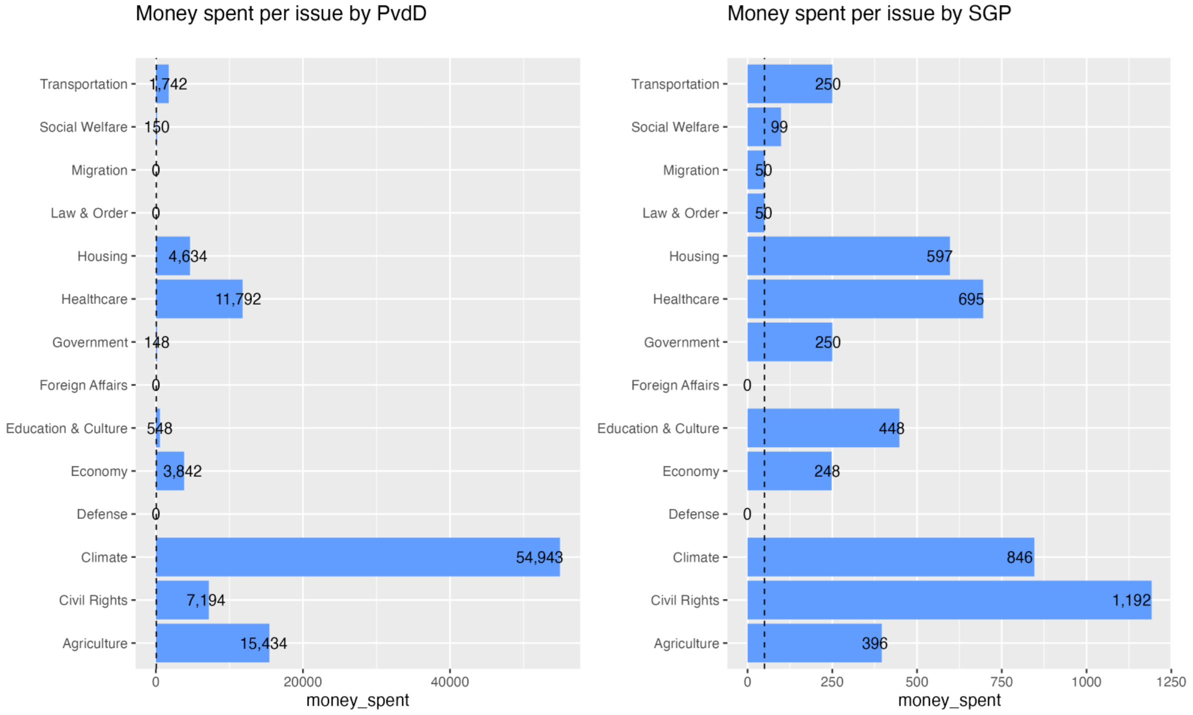


Figure 12. Spending per Issue by PvdD & SGP (in EUR)



https://doi.org/10.33621/jdsr.v7i154885

Figure 13. Spending per Issue by SP & Volt (in EUR)

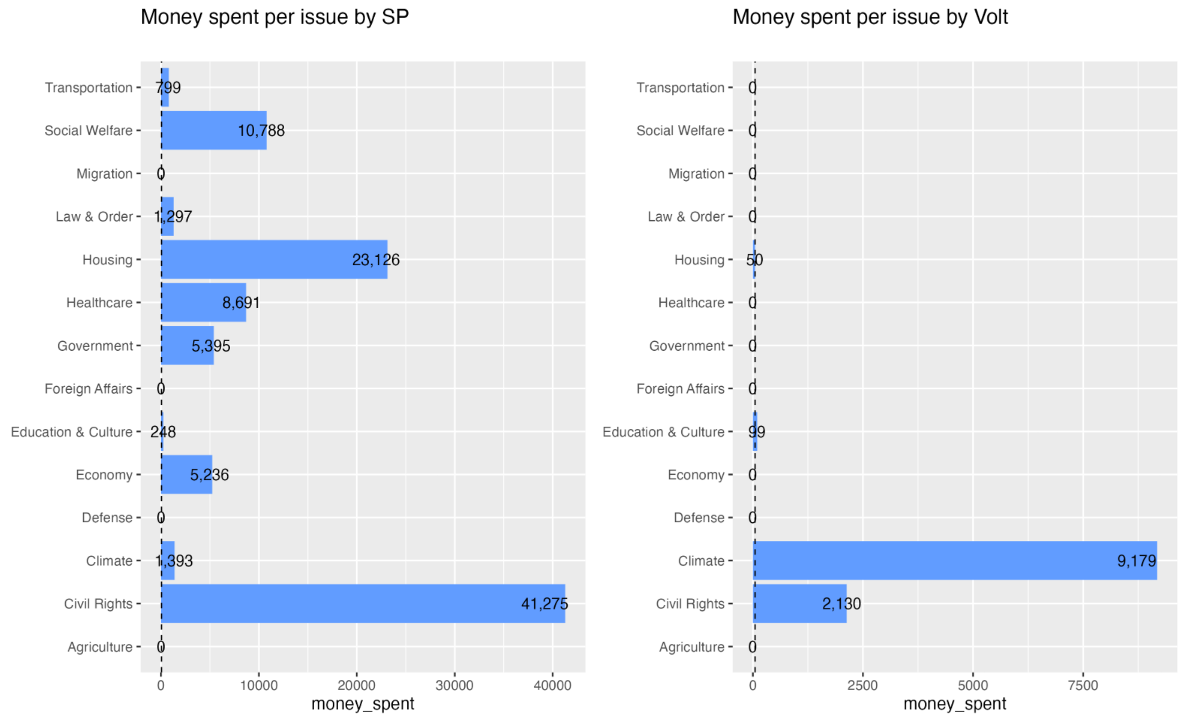
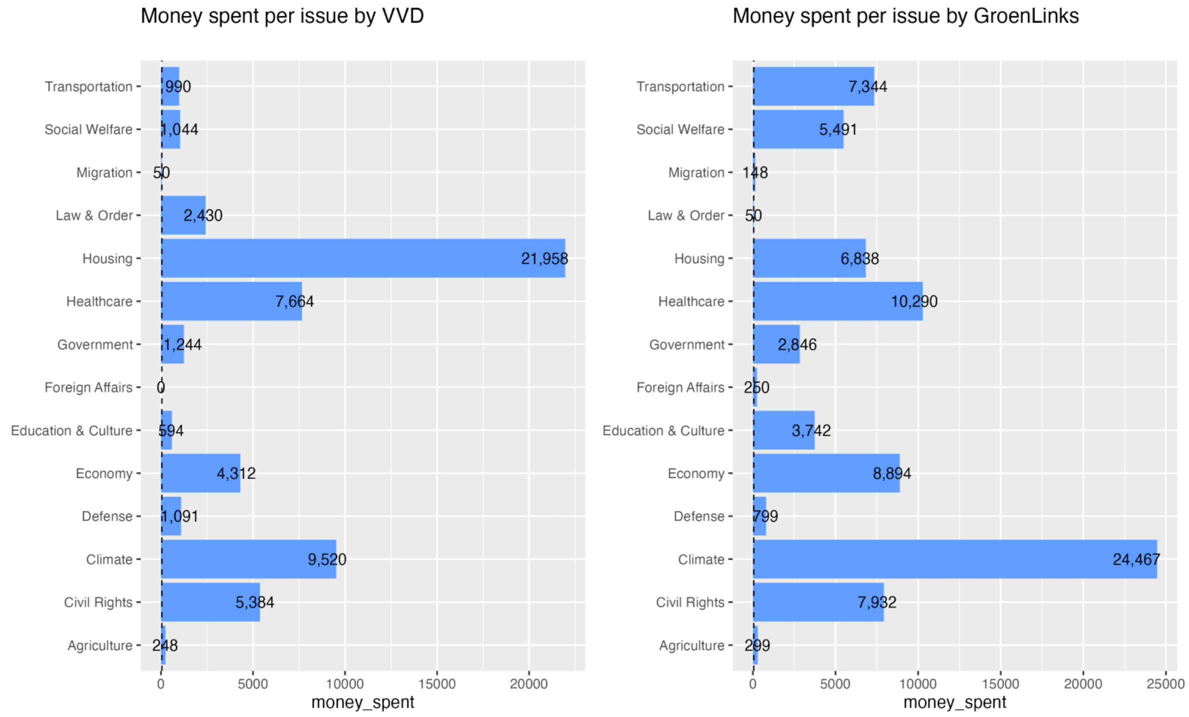


Figure 14. Spending per Issue by VVD & GroenLinks (in EUR)



Appendix F. The number of matched ads per party

Table 7. The number of matched ads per party

Party	# of matched	# of unmatched
50+	96 (59.26%)	66 (40.74%)
BBB	51 (85.00%)	9 (15.00%)
BIJ1	51 (27.42%)	135 (72.58%)
CDA	5,631 (49.12%)	5,832 (50.88%)
CU	83 (36.40%)	145 (63.60%)
D66	558 (36.54%)	969 (63.46%)
DENK	109 (26.33%)	305 (73.67%)
FvD	90 (31.80%)	193 (68.20%)
GroenLinks	174 (39.19%)	270 (60.81%)
JA21	47 (40.87%)	68 (59.13%)
PVV	1 (11.11%)	8 (88.89%)
PvdA	3,455 (78.08%)	970 (21.92%)
PvdD	201 (72.83%)	75 (27.17%)
SGP	54 (51.92%)	50 (48.08%)
SP	165 (38.92%)	259 (61.08%)
VOLT	84 (7.12%)	1,096 (92.88%)
VVD	486 (10.89%)	3,975 (89.11%)
Total	11,336 (44.00%)	14,425 (56.00%)

https://doi.org/10.33621/jdsr.v7i154885

Appendix G. The number of ads per issue per party

Table 8. The number of ads per issue per party

Party	Ag	CV	Cl	De	Ec	EC	FA	Go	He	Ho	LA	Mi	SW	Ta
50+	0	0	0	0	9	0	1	0	18	31	1	0	37	0
BBB	36	3	11	0	4	1	0	2	1	36	1	1	3	3
BIJ1	0	31	3	0	0	3	0	1	6	6	2	0	2	0
CDA	57	1510	130	3	1231	297	0	484	185	823	305	161	1400	11
CU	0	26	13	1	4	12	3	6	11	8	2	2	11	13
D66	1	35	167	0	5	217	0	4	96	41	4	0	13	7
DENK	0	33	0	0	18	32	11	0	12	18	0	6	6	13
FvD	6	11	12	9	7	5	0	3	21	5	3	6	8	3
GL	2	35	66	2	12	17	1	7	20	23	1	3	18	11
JA21	0	0	2	0	11	2	0	0	11	0	7	25	0	0
PVV	0	0	1	0	0	0	0	0	0	0	0	0	0	0
PvdA	3	1500	29	0	255	407	14	19	454	522	7	1	876	14
PvdD	33	12	114	0	17	5	0	3	16	31	0	0	1	15
SGP	8	16	7	0	3	5	0	1	10	6	1	1	2	1
SP	0	50	14	0	27	5	0	10	18	48	6	0	25	2
VOLT	0	39	42	0	0	2	0	0	0	1	0	0	0	0
VVD	3	32	60	18	75	12	0	11	72	183	39	1	13	20
Total	149	3323	671	33	1678	1022	30	551	951	1782	379	207	2415	113