Accessibility statement

This is an accessibility statement for the journal: STS Encounters.

Conformance status

The Web Content Accessibility Guidelines (WCAG) defines requirements for designers and developers to improve accessibility for people with disabilities. It defines three levels of conformance: Level A, Level AA, and Level AAA. This statement is relevant for volume 15, number 1, 2023 and onwards. STS Encounters is partially conformant with WCAG 2.1 level AA. Partially conformant means that some parts of the content do not fully conform to the accessibility standard.

Feedback

We welcome your feedback on the accessibility of the journal. Please let us know if you encounter accessibility barriers. You can reach us at:

E-mail: imvko@cc.au.dk
Address: Helsingforsgade 14, 8200 Aarhus N

STS Encounters
Research papers from DASTS
Volume 18 • Issue 1 • 2026

Stray Data! Stray Publics!
Democratic Interventions with Qualitative Machines

Anders Koed Madsen
Department of Technology, Management and Economics
Technical University of Denmark
https://orcid.org/0000-0003-2842-4908
Inaugural lecture: Department of Culture and Communication, Aalborg University, March 21st, 2025 (in September 2025, Anders took up a professorship at DTU management)

STS Encounters is published by the Danish Association for Science and Technology Studies (DASTS). The aim of the journal is to publish high quality STS research, support collaboration in the Danish STS community and contribute to the recognition of Danish STS nationally and internationally.

ISSN: 1904-4372

Stray Data! Stray Publics!

Democratic Interventions with Qualitative Machines

Anders Koed Madsen

Thank you very much for spending this Friday afternoon with me. It's so nice to see so many of you here, spanning colleagues that I've done research with and taught with, friends and family, but also really importantly for me, those of you working outside university walls that I have done projects with during the last 10-15 years. I'm grateful for you showing up here today. I'm going to talk for around 45 minutes, and my talk is going to be titled 'Stray Data! Stray Publics! – Democratic Interventions with Qualitative Machines'. Hopefully the title of the talk will be clearer to you when we get to the wine at the end of the talk.

I want to start by taking you into a recent conversation with a friend who asked me which academic environment I would like to visit if I had the opportunity to go back in time. I think the answer to that question is actually quite revealing of how you think of yourself as an academic. My choice was New York in the late 30s. More specifically Columbia University and the 'Bureau of Applied Social Research' directed by Paul Lazarsfeld and visited by sociologists like Robert Merton as well as critical scholars like Theodor Adorno. One of the 'applications' developed in this academic environment was the sample survey (Osborne and Rose 1999). With an outset in market studies, Lazarsfeld and colleagues had the ambition of measuring public opinion in more or less the same way that he would measure consumer preferences (Lazarsfeld and Fiske 1938). The underlying ontological premise was that citizens were individuals with internal preferences and values that could be extracted by structured Q&A sessions. Preferably in their own homes without distraction from anybody else. Once extracted, those individuals' opinions could be aggregated up to what would then be termed 'the public opinion'. Gallup later turned this into an industry, which is almost naturalized now as a way of thinking about the public (Gallup and Rae 1940).

Should I ever find myself in 1930's New York, I would also take the opportunity to jump the train to Chicago, and visit 'The Hull House' where Jane Adams (1895) had established a social settlement with a very different way of materializing and assembling a public. Most importantly, Adams' public was an interactive public. The Hull House was designed to facilitate collaborations around art and music and to stimulate debates amongst locals. Importantly, this also involved motivating the inhabitants to map out their own social issues. Inspired by the then recent London poverty maps by Charles Booth (1889), the house produced a series of maps that were supposed to make the social conditions of locals visible to decision makers.

I am inspired by this era of American sociology for two reasons, First, because these attempts at inscribing the public takes place before the divide between qualitative and quantitative methods. Second, because part of the research method was to produce infrastructures for the public and thus intervene in democracy. I will return to that later. For now, I will just note that both these cases illustrate something that Monique Girard and David Stark has formulated quite nicely, namely that:

> "[T]here is no public, no public assembly, without protocols and technologies – even if these are as simple as chairs around a table and everyday conventions of conversational turn-taking" (Girard and Stark 2007).

Personally, I got first-hand experience with such protocols and technologies early in my bachelor studies, where I got a student assistantship at the Danish Board of Technology. Most of you probably know the board, but for those of you who don't it has been pioneering in inventing formats for deliberative democracy such as the consensus conference (Joss and Durant 1995). As a young student assistant, I found myself moving chairs around, discussing how many people should sit around a table, how long each person should be allowed to speak, and through which procedures it was fair to cluster the post-its left by participants. We were negotiating small practical details that were very nonetheless important parts of making a public visible to the Danish parliament at that point in time.

Drawing on anthropological theory, one could say that we tried to build a home for the public (Pinel and Svendsen 2024). We tried to establish some temporal and physical boundaries where public assemblage could take place, the norms and rules for the interaction to occur in that time-space and the way such interactions would be condensed into a form of legitimate consensus. We developed strategies for domesticating the public and its empirical traces. However, the caveat with this strategy is that every time you build a home, somebody always wants to leave. In the context of participatory democracy people have always taken advantage of available communication channels to organize themselves as a public outside the scope of the more established frames.

Throughout history, different labels have been put on these unorthodox publics. Blumer (1948) talked about them as 'acting publics', Michael Warner (2021) proposed 'counter publics' as a fitting headline and Noortje Marres (2005) denoted them 'issue publics'. There are slight differences in these different conceptualizations, but they share one important trait. Compared to most deliberative theories, they lower the barrier for what it takes to be part of a legitimate public. For instance, Warner explicitly argues that maybe sometimes paying attention to something is enough to be part of a public. The decision to engage in so-called rational debate is just one of many ways to exercise oneself as part of a public. Today, I prefer to use the term 'stray publics' to signal that these publics act outside of the institutionalized homes built for them. They are not in the citizen hall or at the receiving end of a telephone line answering structured questions from a Gallup employee. They roam on their own and make up their own rules for engagement.

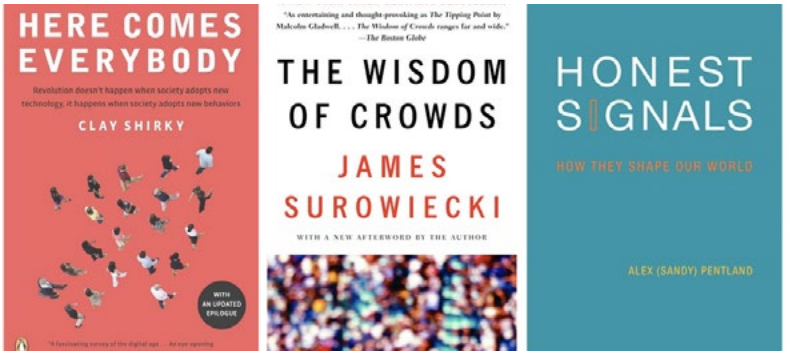
The result is that they are very hard to inscribe and quantify. Compared to domesticated publics they leave 'stray data' in a form that is often hard to recognize as data at all (Madsen 2018a). Whereas Likert-scales and – to some extent, post-its at the city hall – have been legitimized as public signals, the same is not true for graffiti, banners, tweets, likes and other forms of by-products of the various kinds of publics that stray from their institutionalized homes. Because they form organically around infrastructures at their disposal (urban furniture, digital platforms, etc.) they are hard to inscribe. This is something that was already noted by Herbert Blumer (1948) in his 1948 address to the American Sociological Association, where he lamented the fact that methodological individualists like Lazarsfeld had an easier time operationalizing their imaginary of the public than interactionist scholars interested in processual and organic publics. As he put it, "the problem of where to dip in, how to dip in, and how far to dip in is what I have in mind when speaking about sampling an organic structure"

During the time I worked at The Danish Board of Technology, we were confronted with a very specific form of stray public that also manifested itself in forms of data that were a bad fit in institutionalized public engagement toolkits. Here I am speaking of the digital publics that formed around the then new platforms of the participatory web in the mid 2000s (Bruns 2008; O'Reilly 2009). I worked at the board from 2005–2008, and around that time we saw the emergence of LinkedIN (2003), Facebook (2004), Twitter (2005), Reddit (2006), newspaper commentary sections such as Ekstra Bladets 'Nationen' (2007), Youtube (2007) and Instagram (2009).

As the iPhone (2008) suddenly put all these platforms in the pocket of the user we were faced with an collective of people were suddenly expressing their views without necessarily showing up at the dedicated homes we tried to built for the public.

It's no wonder that we asked ourselves questions similar to those Blumer had asked 60 years before. Where and how to dip in? Whereas conventions for sampling individuals in representative manners existed, there was no such conventions for the digital publics that roamed around us. One obvious response was to deem these publics illegitimate on the notion that they did not engage in proper democratic activity because their interactions and discussions failed to live up to the criteria for rational dialog enforced in the physical meetings. For instance, when evaluated with an outset in Habermasian theories of deliberative democracy, the emerging online spaces did not constitute a proper 'democratic situation' (Papazu and Birkbak 2022).

However, the 'zeitgeist' in the mid–00's was not exactly Habermasian. After Google's success in building a functional search engine on top of 'exhaust data' – such as hyperlinks – in the late 1990's, social theorists – especially in the United States – began to explore the democratic potentials in learning from the data people left as part of their everyday lives. When I started my MA in Chicago in 2008, the books we read in the first semester bore titles such as "The Wisdom of Crowds" (Surowiecki 2005), "Here Comes Everybody – the power of organizing without organizations" (Shirky 2009) and "Honest Signals" (Pentland 2010). It is hard not to detect a certain sense of tech-optimism here. First, some of them promoted the idea that the data people leave as part of their everyday life, are in some sense more honest than the ones they leave as respondents to top-down processes such as polls and citizen hall meetings. Second, all of them promoted some version of the idea that algorithmic techniques would be capable of finding meaningful patterns in the messy data left by people online. A hope that perhaps reached its most extreme formulations in Chris Anderson's (2008) infamous essay proclaiming a shift from theorizing to data-mining in the sciences and O'Reilly's (2011) dream of algorithmic governance as the future for the American bureaucracy. On the more practical side initiatives such as Google Flu Trends (2008) served as one of the demonstration cases because it – for a while – succeeded in predicting flu epidemics with an outset in semantic patterns in search terms from American Google users. Being able to detect flu outbreaks faster than statisticians in Center for Disease Control seemingly demonstrated the potential in learning from exhaust data (Lazer et al. 2014).

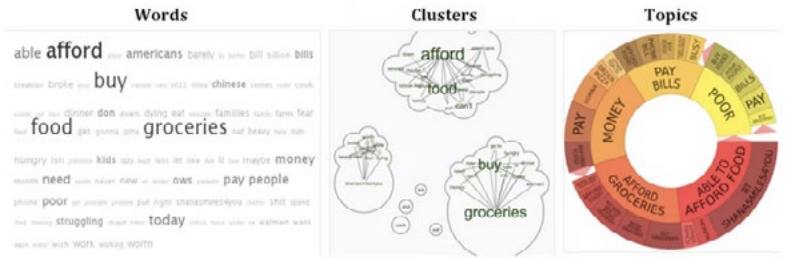


It is interesting to note that most of the literature and demonstration cases had their origins in American social theory. A tradition that has never drawn as sharp a theoretical distinction between 'crowds' and 'publics' as was the case in Europe (Borch 2013). This is exemplified in the writings on both sides of the pond at the historical epoch that I opened the lecture with. Whereas European scholars such as Gustave le Bon and Gabriel Tarde took crowds and crowd psychology as an antithesis to democracy and public engagement, this was not the case for their American counterparts such as Robert Park. Whereas European writers used medical metaphors such as 'contagion', 'hypnosis', 'delirium' to describe the psychological status of the crowd, American writers at large took a different normative position. For instance, Park suggested that crowds and publics are two variations of the same social form - both have the capacity to 'bring individuals out of old ties and into new ones' (Park as quoted in Borch, 2013, 14). On this view, crowds have liberating potential with clear democratic value. I mention this because I got my postgraduate training in the USA.

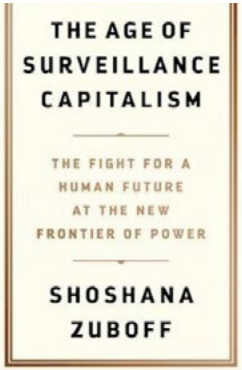
Being brough up on American pragmatism and has undeniably shaped my thinking on the relation between crowds and publics. I continue to believe in the democratic potential in listening to stray publics through algorithmic analysis of their exhaust data.

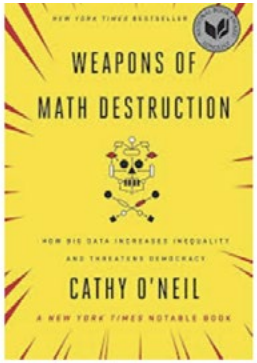
But…and here the 'but' is really important. Entertaining this belief does not entail subscribing to the naive empiricist hopes voiced by the above-mentioned scholars of the 2000's. The dream of a science without theory, of a mode of organizing without organizations, and of the existence of honest data signals is exactly that – a dream. It should not be the epistemological ground from which we assess the democratic value of crowd data. Formulating alternative grounds – and accordingly different analytical techniques - is one of the main academic challenges that I have grappled with the last decade. For me, this journey towards 'another' ground took shape during my PhD studies at the Department of Organization at Copenhagen Business School where I had Maja Horst as my supervisor. Being embedded in an academic community rounded by Science and technology studies (STS), valuation studies, and related branches of organizational theory taught me a key lesson. Namely, that data is not convincing because of its epistemic qualities and predictive performance alone. In order to serve as actionable evidence data needs to be networked into pre-existing epistemic cultures, professional visions and notions of expertise.

This was also something I observed first-hand when studying data work in the United Nations sub-branch 'Global Pulse' during my PhD studies (Madsen 2015). Established as a data innovation hub, the Pulse employees had the ambition of using data from Twitter to predict economic crises around the world. The idea was that since people suddenly had handheld telephones, they would be able to tweet about food prices when they went to the market. And, actually, people did that. So, the Global Pulse built a dashboard that could be used to assess whether there was an emerging economic crisis much faster than the household surveys that preceded it. Almost replicating what Google Flu Trends had done for epidemiology in the context of foreign aid. The interesting thing was that while the model had a good predictive performance as seen from a technical perspective, it wasn't actionable within the epistemic culture of the UN for two reasons. First, because the involved digital methods could not be standardized in the way that the UN requires. Twitter as a platform changes all the time and therefore it's hard to develop a standardized, stabilized framework or method for working with tweets that can be approved by the organization. Secondly, because the culture of using Twitter differed from country to country. The Pulse employees would have to invent a new baseline for every area in order to color the dashboard red or green. The data simply didn't fit the organizational structure and its expectations for evidence. The translation from dream to practice was not as smooth as imagined. A reality that also caught up with Google Flu Trends that made headlines in 2013 for overestimating the flu partly due to the addition of auto-complete in Google's search bar (Lazer et al. 2014).



While these troubles gradually became apparent to 'big data' practitioners in the 2010's, this same decade also saw a normative 'vibe shift' in social theory on data. Suddenly the best-selling and most cited books bore titles such as "The Filter Bubble – What the internet is hiding from you" (Pariser 2011), "The Age of Surveillance capitalism" (Zuboff 2019.) and even "Weapons of Math Destruction – how big data increases inequality and threatens democracy" (O'Neil 2016). In contrast to the books a decade earlier, exhaust data and algorithmic pattern recognition was suddenly positioned as an antithesis to democratic life. Rather than a source for listening to distributed public data, algorithms were largely framed as techniques of control and manipulation.



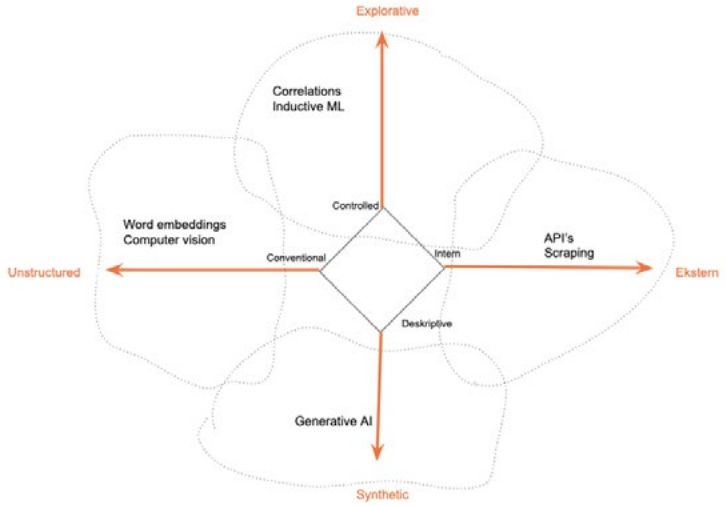


TANTLab: An experimentarium for qualitative machines

This was the moment when I entered the techno-anthropological laboratory (TANTLab), where I have been working for more than 10 years now. It was also around the same time that I co-founded the Public Data Lab with a handful of European colleagues working at the intersection of data science and SSH. What characterizes both these institutions is that the involved researchers have made efforts to balance the potentials in listening to stray publics through their exhaust data while at the same time acknowledging the pitfalls involved in that endavour. On the one hand, we have been driven by a fascination about the potentials to 'sample organic publics' with digital methods. On the other, we have kept a strong theoretical focus on the performativity and unintended consequences of such new democratic data practices. In TANTLab we have done this work at the intersection of digital methods, science and technology studies, anthropology of technology, and interventionist research in quite specific ways. We have made it an ethos to study the democratic impacts of algorithmic knowledge production by practicing it. Often through 'data-sprints' (Munk, Meunier, and Venturini 2019) with external organizations interested in experimenting with measuring social phenomena in new ways and being open about the extent to which assumptions about the value of computational methods were – or were not – met in the process.

I want to exemplify this approach with a project that turns 10 years old this year. It is a project that I did with Anders Kristian Munk, Thøger Riis Michelsen and others and it took its outset in the school reform of 2012 (Madsen and Munk 2019). A reform that left municipalities with the task of implementing a new school system in dialogue with key stakeholders. In the municipality of Aalborg a decision was made to run a participatory process. However, it turned out that those with a stake in the school were more active on Facebook discussion threads than in the formal political meetings. As the relevant public 'strayed', the municipality was left with a decision to either deem the online public illegitimate or develop methods for listening to it. They chose the latter, and we were brought in to invent a way to visualize the debate and feed it into the political budget prioritizations. One interesting unintended consequence of this democratic intervention was that the visualization we produced ultimately served to shortcut the Teachers Union's ability to act as spokesperson for a unified group of teachers. As the teachers contributing to the debate were visualized as individuals with their own distinct take on the reform, they seemed less in sync as an interest than they would otherwise appear. A practical consequence that also made us revise our own assumptions about the democratic value of our intervention.

During the last decade we have continued this kind of data experimentation – not least with the students on our education in data-driven organizational development. In order to explain the character of these experiments, the 'the data-diamond' – illustrated below - comes in handy (Madsen, Søltoft & Klitgaard, 2025). It illustrates how most of our data experiments have started from a situation where the data habits of our collaborators are placed in the small diamond in the middle. They typically work with data that is quite conventional. It comes in numbers in an Excel sheet, and it is typically produced within the organization or produced by a trusted partner. The analysis of that data is controlled in the sense that it is driven by the desire to disprove a null-hypothesis. Finally, it is communicated in standardized descriptive genres such as bar charts and line graphs. The experimentation we engage in consists in moving the data practices along the axes and observe the consequences of doing so. On the vertical axis this involved experimenting with new data types such as unstructured data sources or data produced by actors that are out of the control of the organization. On the horizontal axis it involves experimenting with new modes of data analysis such as explorative statistics and synthetic data.



In a moment I will exemplify this research strategy with my current research on digital placemaking and urban planning. For now, I just want to say that what interests me is that movement along these axes opens for practicing versions of data science that are more in tune with qualitative epistemologies than classic statistics. It affords working with 'qualitative machines.' For instance, combinations of unstructured and explorative techniques make it possible to engage in a form of distant reading (Moretti 2013) where data science supports the practice of seeing patterns in texts or images and then do close reading of strategic samples. Inductive machine learning such as clustering techniques opens for working with situated classifications in a more ideographic style and the possibility to work with data from various API's carry the potential to model emic perspectives with an outset in people's own language and conceptual systems (Nelson 2020; Munk and Winthereik 2022). Something that can even be manifested in synthetic agents carrying such perspectives through generative AI (Nelson forthcoming 2026; Pilati, Munk, and Venturini 2024). I label these forms of machine learning as qualitative because they are best approached with an outset in epistemological traditions from disciplines such as ethnography, media studies, history and literary studies. Also, they lend themselves to very different ways of interfacing with data than the standardized techniques assembled in quantitative software. An example is the possibility to chat with data instead of visualizing it in Power BI and Tableau. This opens for new data imaginaries that are much more akin to the epistemologies of the humanities even though the underlying data is ultimately quantitative.

Soft city sensing – bringing the TANTLAB ethos to urban planning

This brings me to my current research which revolves around data experiments in the context of urban planning. During the last eight years I have worked with architects, municipalities, DMO's and urban planners on moving their data practices along the above-mentioned

axes to test the potentials of knowing the city through qualitative machines. I am currently writing this up in a book called 'Soft City Sensing – Listening to Data Publics with Qualitative Machines' (rooted in Madsen, Grundtvig, and Thorsen 2022). Hopefully that title makes sense now that we've been through the first part of the talk. The main aim of the book is to provide an alternative to dominant imaginaries around data-driven urbanism, which is usually associated with the form of smart city control rooms seen to the right below (Madsen 2018b). A room where urban decisionmakers can measure and follow how urban things – cars, people, trash – move in physical space. It is a data imaginary driven by the Internet of Things (IoT)and the ambition of those projects is typically to make the city more efficient and increase the predictive power of professional planners (Kitchin, Lauriault, and McArdle 2015). As such it is a mode of thinking about urban data that can be traced all the way back to the physio-economic simulations in the 19th century (Batty 2024).

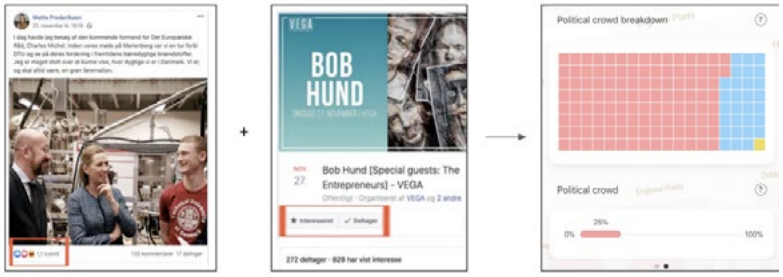
In contrast, the kind of urban data science I propose in the book is more like a computational version of traditions such as Milgram & Jodelet's (2014) cognitive mappings of Paris portrayed next to the Copenhagen dashboard above or the descriptive regional geography on the 1930's (Hartshorne 1939). Building on these traditions, the core of the book is to formulate a theoretical, philosophical and methodological backdrop for using exhaust urban data and qualitative machine learning to develop cities - while also attending to the various pitfalls in doing so. As people increasingly engage with cities through digital platforms (Halegoua 2020), the data from these platforms can provide insights into the softer aspects of urban life that is not just about efficient flows of things. In its more case-based part, the book has four empirical focus area to illustrate this potential. First, it attends to the possibility of using online traces to inquire into the way people remember the past and the timescapes they mobilize when enacting this memory (Kitchin 2023). A particularly valuable source for this is the many Facebook groups revolving around sharing personal urban history. Second, it attends to the possibility to understand people's perceptions of urban atmospheres and aesthetics through sharing images and sound (De Matteis et al. 2019). Third, it will most likely have a section on tourism-related issues and fourth, it has a section on social infrastructure, which is the empirical theme I want to unpack a little bit now.

Social infrastructure, as I take it, concerns the ability of the city to bring people together across differences and create a sense of belonging for people living in the city or working in the city or just visiting the city (Klinenberg 2018). It concerns the role of the city in building social connections. One project that I discuss in this part of the book is a project called 'Do You Live in A Bubble' (Madsen 2022). In collaboration with GEHL architects and colleagues from TANTLab[1], the project set out to map the political diversity of Copenhagen. Our guiding ideographic question was whether Copenhagen succeeds in creating places where people with diverging political views meet. A question rooted in the normative assumption that part of what makes a city democratically valuable is its ability to bring people together despite differences. Something that major cities – especially in the USA – tend to fail on (Brown and Enos 2021). In our initial research we found that the most prevalent way of operationalizing this problem was to plot voting statistics on geographical voting districts. To map whether people who live together vote in the same way. A way of operationalizing the problem that is positioned right in the middle of the data diamond above.

[1] Done together with the Facebook Atlas project

The problem with that operationalization was that it did not really capture people's urban life as we wanted to capture it. Most importantly, it tied people to their residential house or apartment. It did not provide information about the way they engage with the city besides the information on where they sleep. Also, the fact that voting data comes every second year and aggregates into pre-defined voting districts means that it is not very granular, neither in space nor time. As an alternative, we decided to work with Facebook data for the simple reason that Facebook is a platform where people leave traces of both their political and urban preferences. Our operationalization of the political diversity of Copenhagen was the following. First, we provided 300.000 anonymous users with a metadata on their political leaning based on their interactions with content on dedicated political Facebook pages. If an ID was predominantly linking or hearting content from the Social Democratic Party, SF, The Alternative or The Red/Green Alliance they would be seen as leaning towards the 'red' block in Danish politics. A similar strategy was used to position users as leaning towards the 'blue' block (economic liberals) or the 'yellow block' (nationalists). Second, we followed how these 300,000 users interacted with 150,000 events in Copenhagen. For instance, if a 'red' user would indicate that he/she was interested in – or planned to participate – in an event, that event would get a 'red point'. Looping over all users and events enabled us to do a detailed political breakdown for each event.

Now, one objection from the perspective of people working on deliberative democracy would be that these are very 'thin' traces. Being left on a commercial platform driven by algorithmic recommendation systems, it is fair to discuss whether this is a legitimate trace of a public. Drawing on my arguments in the opening section, I want to make the argument that it really is. An argument that follows in line with Michael Warner's argument that we put too much emphasis on dialogue and deliberation when we assess the value of inputs from the public. In fact, Warner suggests that sometimes signaling attention is enough of a public act to be taken seriously. As likes distributes attention they fall within this category and – in this specific project – we ultimately used them to build an interactive map of Copenhagen's political diversity. It is published online and allows the user to zoom in and explore the political profile of most places and areas in the city.



Vague situations and reflexive models

So, what is the interesting thing about a map like this one? One thing is for sure; we cannot use it to steer mobility in the same way as the IoT-infused control room above. As we do not know whether people actually visited these events physically, we do not have a backing to infer behavioral patterns. What we can know from this data is mainly how people have chosen to publicly signal their associations to places and politics. While likes on events do not mean attendance, they do carry useful information about patterns in people's urban preferences and self-presentation. Contrary to the data optimists in the 2000's, I do not think of this information as more honest or neutral than any other

data type. Also, I am not of the opinion that 'big data' trumps small data. The map above is not the final evidence on the problem of political diversity. But it is valuable because it provides new clues to places in the city worthy of qualitative deep dives as well as new frames within which the problem of political diversity can be formulated.

The map carries democratic and epistemic value because it produces what John Dewey (1929, 1910) would describe as a vague empirical situation. A situation characterized by the fact that you are not quite sure how to describe it. Which empirical objects are present? How do we classify them? How do they relate to each other? What is valuable to focus on? Because these fundamental frames are unsettled, vague situations are productive for learning. In fact, Dewey makes the point that the way we learn and develop as humans and organizations is precisely by positioning ourselves in problematic situations where we're unsure about what it is that we are looking at. Furthermore, he saw a strong connection between productive learning environments and productive democratic situations. In many ways his theory of democracy was a theory of (co)-learning and this is the lens through which I will justify the value of the kinds of maps produced above (Madsen 2024). Because the underlying data is unlike the data usually used to frame problems of urban social infrastructure, it opens for reframing this problem in ways that can both stimulate insight about the city and perspectives on where it carried democratic value.

In this specific project one frame that is troubled is the so-called neighborhood unit. Ever since Clarence Perry (2013) popularized this concept 100 years ago, the idea that urban social infrastructure is produced by getting people to share residential locations has been very strong (see e.g. Cox 1969; Duncan and Duncan 1955). The assumption has been that if people live in geographical proximity to each other, they will also establish social proximity to each other. They will create a community. If a community is coupled to neighborhoods in that way, then it follows that neighborhood design – with its distribution of libraries, schools, affordable houses and similar public amenities – becomes a form of social engineering. While this is, of course, an important part of planning for a mixed city, the map above nonetheless provides an alternative story.

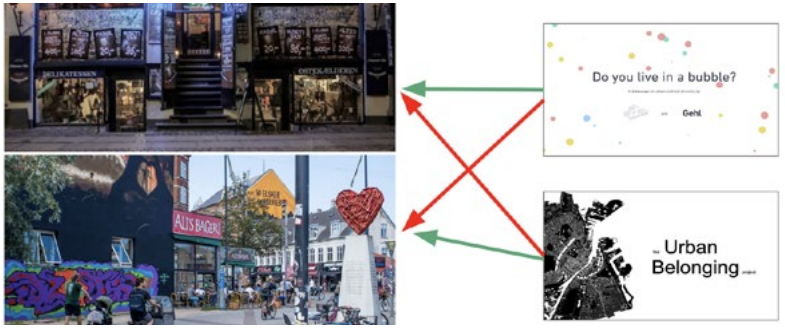
For instance, many of the areas that are geographically approximate are actually very different when it comes to political diversity. One example is the Copenhagen peninsula called Refshaleøen seen below. One part of the island caters for a specific 'red' crowd, whereas only the eastern part of the island caters for a politically diverse crowd. If we zoom in and take a qualitative look at what goes on in this eastern part, the urban life here is characterized by people who meet to do BMX or climb together. Because the data is relational (we can follow a user through various events), we can also begin to trace whether or not the politically diverse crowds meeting and climbing at Refshaleøen have equivalences in other parts of Copenhagen. From the map it seems that activity in other parks such as Valby Parken and Fælledparken carries similar political profiles despite their low geographical proximity.

This, of course, raises the question of how to design for political diversity in the city. Do we need professional designers to plan the physical infrastructure or can we also achieve diversity by conserving the activities that already perform what we want to achieve? Perhaps the answer to the politically diverse city requires a typology of activities to support rather than a blueprint for good neighbourhood design. If that is the case, then the notion of who is the expert in designing the diverse city is certainly an open question.

Maybe it's the guy who owns the climbing hall who really knows about the design of diverse spaces. This proposition is one productive conceptual frictions emerging from the map.

But the map stimulates productive normative friction as well. In the project it ended up troubling some engrained ideas about what a good city is. Below, you see two places that appeared as highly politically diverse. One is the motorcycle repair shop in the meatpacking district. Another is the fans meeting at Parken's football stadium to organize the coordinated visual spectacles known as 'tifos'. What is interesting is that while these places came out as having democratic value on our diversity metrics, they both fare quite badly on other urban metrics used to measure the desirable city. For instance, if we measure the city as a particle emission zone, we would be ill advised to place motorcycle repair shops in the center of the city. The bikes would have to ride to the periphery afterwards and carry a trail of pollution with them. Similarly, the football fans who seem to play a useful democratic role have various types of negative effects, such as draining the police budget and creating a space that is not very welcoming to women. As such, the map produces valuation conflicts that were not visible before the project started (Kornberger et al. 2015).

The fact that 'soft city sensing' projects can stimulate normative frictions becomes even more apparent when looking at the political diversity project in combination with another project, called Urban Belonging (Madsen et al. 2023; Burgos-Thorsen, Niederer, and Madsen 2024). In this project, the data strategy was slightly different. Rather than obtaining exhaust data from online platforms, we developed a photo voice app that allows people to take pictures of the city and tag them based on the story they want to tell. More specifically, we mobilized marginalized groups in Copenhagen – homeless people, LGBTQ environment, ethnic minorities etc. – to tell stories about their belonging to the city. What came out of that project was mental maps of different groups showing how they moved about in the city, portraying it in pictures. The interesting finding in terms of normative friction is that when you compare these two projects, some of the same places come out with very different evaluations in terms of their social qualities. A good example is the street of Gothersgade in the inner city of Copenhagen. It is highly politically diverse, but at the same time a place where many of the participants in the urban belonging project did not feel safe. Primarily because it is a street that is dominated by a specific, perhaps masculine, energy. To the contrary, Nørrebro was a place where many in the urban belonging project expressed a feeling of safety and belonging. But it is also a very politically homogeneous area. In fact, people from the 'yellow' block have expressed that they do not always like to reveal their political preferences when they go out there. It is definitely not seen as a safe space for them.



These kinds of frictions are of democratic relevance because they raise interesting questions concerning how we think about urban design and expertise. For instance, the juxtaposition of these two projects indicates that the ideal of universal design – the hope that it is possible to design all-inclusive spaces – is perhaps not always feasible. One suggestion from the juxtaposition is that places feel inclusive to some people precisely because they leave others out. Nørrebro is the perfect case here. It is a safe space for some groups because other groups are not welcome or nudged to stay silent. Following this reasoning, maybe, the inclusive city is a conglomerate of partially inclusive spaces with no space being universally inclusive. This is an interesting proposition because it troubles some of the powerful narratives in contemporary urban planning. While we have for a long time known that we should not design for the average man, maybe the solution is also not to design from marginalized perspectives and then expect that they will work for all. From the two projects we can see that the marginalized do not function like a least likely case in this way.

To reiterate, the value of experimenting with qualitative machine learning in the ways discussed above is that they produce useful frictions with established data practices (Madsen, Munk, & Søltoft 2023). In that sense they open a democratic situation where social problems can be reframed and new forms of expertise can gain legitimacy. In my future research, I would like to push this agenda further by taking inspiration from discussions of 'reflexive modeling' in organizational theory (Beunza and Stark 2012b, 2012a). Under this heading we have seen fantastic ethnographic studies of how financial traders, for instance, interpret and use their financial models. The finding is that few of them trust models as representations of the market one-to-one. Rather, they cherish their models because they make different valuation regimes visible. While each model carries its own biases, the possibility to juxtapose a variety of models built from different assumptions is what traders find valuable. Models are cues to how other traders think. Rather than giving answers they are heuristics for critical reflexions at specific moments in the process of arriving at a trading decision. On this account reflexivity in financial trading is not a mental process or solipsistic practice, but it springs out of a process of distributed cognition involving the strategic contraposition of different material artifacts.

This way of thinking about the role of computational modeling is relevant to explore further in the context of what has been discussed above. Even more so in the context of synthetic agents on the rise as new forms of computational methods (Søltoft, Kocksch, and Munk 2024; Madsen & Søltoft 2023). One of the interesting potentials of Large Language Models, and some of the generative AI that we are confronted with right now, is exactly that they are biased and carry cultural perspectives through their training. They open for an approach to data science that Laura Nelson has dubbed 'perspectival modeling' (Nelson forthcoming 2026). It involves training models to imitate specific societal perspectives and thus enabling them to travel and enter into communicative practices in new ways (Esposito 2022). This is an interesting prospect to explore further in the context of organizational learning. Perhaps the possibility to interface with synthetic versions of diverse stakeholders offer organizations and decision-makers and escape the form of groupthink that is a constant risk if one strategizes in a context of low cognitive diversity. On this account 'synthetic' is not in opposition to 'authentic'. It is rather understood in the Kantian sense where synthesis is a necessary condition for ordering the world and making empirical knowledge claims (Kant 1787). The prospect of synthetic agents is an opening for experimenting modes of synthesis that are foreign to one's own conceptual routines. This is something that colleagues and I recently explored in a project with Henning Larsen Architects where we used generative AI to create images of the future of Refhshaleøen based on the pictures that people had taken out there on Instagram during the last 10 years. Going back to the title of this talk one could say that the idea was to feed data from stray publics into a generative machine and then project their perspectives into renderings of the future.

Computational SSH

These types of experiments with computational methods lead me to the last topic I want to cover in this talk which is the part of my professorship that is directly focused on developing Aalborg University's profile in computational social science and humanities. Part of my employment is to serve as scientific director of MASSHINE – the dean's signature project for computational SSH. 'SSH' is put in the middle here because we want to signal that our intention is to integrate epistemologies from social sciences and humanities into machine learning techniques. We want to design computational methods with an outset in the traditions of SSH disciplines rather than importing them from the outside. I want end my talk here by outlining three visions or strategic goals for this initiative in the coming years.

The first touches on the identity and self-image of the many SSH disciplines and researchers that define themselves in opposition to quantitative methods. With the rise of qualitative machines such as the ones discussed above, I think we need to overcome the divide between qualitative and quantitative methods. At least if this divide is taken to pertain to the types of data we are working with. The reason being that most digital data is both qualitative and quantitative in nature (Moats 2021). Let us take digital images on Instagram as an example. On the one hand they are qualitative. They derive from an intention to tell a story, and they are produced in a specific context that needs to be understood to gauge its meaning. On the other hand, such images are also a collection of pixels that can be measured in terms of color and composition (Niederer and Colombo 2024). I know that this has been said for at least 15 years, but we need ways of talking about methodological traditions without breaking them into 'qual and quant' from the outset. Perhaps focusing more on differences in approaches to data analysis than types of data. For instance, one can work with quali/quant data with an outset on both nomothetic and idiographic approaches to analysis (Jacomy 2020). In continuation of that, I also think we need to revisit the competence profiles of humanities and the technical sciences. In research collaborations around machine learning and AI it is oftentimes the case that SSH researchers get typecasted as the domain experts that can get methodological help with new methods from computer scientists. On this account, SSH has questions and theories, whereas computer science has methods. However, in a world of qualitative machines this engrained division of academic labor might not make sense. If most of the world's data come in unstructured forms such as texts, sound and images there is no reason why SSH researchers should not position themselves as experts on data methods as well. Especially in a situation where vibe coding promises to lower the technical barriers for writing code and building models.

These points lead nicely to thoughts about the outputs of future SSH research. To paraphrase my colleague, Mathieu Jacomy, I believe that we should start valuing toolmaking and tool maintenance much more. Looking at the way our research is evaluated at SSH it is fair to say that we still have a tendency to see ourselves as a text producing business. We value the publications we do and the citations they get, but we do not in the same way value publishing code on GitHub and sharing it with others. In my opinion we need to start doing that because it is the only way to stimulate the production of the type of SSH-inspired machine learning that I have suggested throughout this talk. Building tools from within SSH will also have didactic potential when teaching the potentials and pitfalls of AI to SSH students. It is crucial that our students get a certain 'critical proximity' (Latour 2005; Birkbak et al. 2015) to the technologies they end up using or criticizing. The worst move we can do as SSH researchers is to allow ourselves the privilege to criticize computation at a distance without engaging in concrete experimentation ourselves. The fact that such experimentation and toolmaking can also lead to high quality critical research was illustrated twelve years ago. Engaged in building data collection for media research digital scholars like Bernhard Rieder were amongst the first to notice the loopholes in Facebook's API that later became the root of the Cambridge Analytica scandal (Rieder 2013). In my role as scientific

director of MASSHINE I will push for an infrastructure that makes this experimentation available for scholars and students alike.

Finally, I believe that scholars in computational SSH have an important societal task in front of us when it comes to lobbying for data access and formulating ethical guidelines for working with this data. After the so-called 'APIcalypse' (Bruns 2021) data access has become hard and it seems that big tech has to some extent misused data protection legislations to also establish themselves as data monopolists. This leaves us with hard legal and ethical questions concerning our right – and perhaps even responsibility – to find ways of analyzing some of the data processes that are central to the functioning of our society. For instance in relation to the spread of misinformation on social media platforms during elections. Guidelines on how to deal with the current data situation cannot be formulated in isolation by initiatives such as MaSSHINE. We need to collaborate across the Danish initiatives that are working on similar agendas within the social sciences and humanities. As well as with international peer communities such as Association of Internet Researchers (AoIR) which has been vital in formulating research guidelines on these matters (Ess 2017). I see it as part of my role as scientific director of MASSHINE is also to help build these relations.

Bio

Anders Koed Madsen is Professor of Data, Democracy, and Digital Urbanism at the STS section at Technical University of Denmark. His research examines how machine learning applied to large-scale qualitative data can inform policy-making and urban planning in ways that respond to local challenges. Under the concept of *Soft City Sensing*, he has collaborated with urban design firms such as Gehl, Henning Larsen, and BIG. His work combines pragmatist philosophy, computational humanities, internet studies, and organizational analysis.

Anders is co-founder of the Public Data Lab, has contributed to networks advancing humanistic approaches to data science, and held visiting positions at Columbia University, Goldsmiths, and MIT.

References

Addams, J. (1895). *Hull-House Maps and Papers: A Presentation of Nationalities and Wages in a Congested District of Chicago*. University of Illinois Press.

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired Magazine*, *16*(7), 16–07.

Batty, M. (2024). *The Computable City: Histories, Technologies, Stories, Predictions*. MIT Press.

Beunza, D., & Stark, D. (2012a). From dissonance to resonance: Cognitive interdependence in quantitative finance. *Economy and Society*, *41*(3), 383–417.

Beunza, D., & Stark, D. (2012b). Seeing through the eyes of others: Dissonance within and across trading rooms. In A. P. Karin Knorr Cetina (Ed.), *The Oxford Handbook of the Sociology of Finance*. Oxford University Press.

Birkbak, A., Petersen, M. K., Elgaard Jensen, T., & Philosophy Documentation Center. (2015). Critical proximity as a methodological move in Techno-Anthropology. *Techné: Research in Philosophy and Technology*, *19*(2), 266–290.

Blumer, H. (1948). Public opinion and public opinion polling. *American Sociological Review*, *13*(5), 542–549.

Booth, C. (1889). *Labour and Life of the People*. Williams and Norgate.

Borch, C. (2013). *The politics of crowds*. Cambridge University Press.

Brown, J. R., & Enos, R. D. (2021). The measurement of partisan sorting for 180 million voters. *Nature Human Behaviour*, *5*(8), 998–1008.

Bruns, A. (2008). *Blogs, wikipedia, second life, and beyond: From production to produsage*. Peter Lang.

Bruns, A. (2021). After the "APIcalypse": social media platforms and their fight against critical scholarly research. In *Disinformation and Data Lockdown on Social Platforms* (pp. 14–36). Routledge.

Burgos-Thorsen, S., Niederer, S., & Madsen, A. K. (2024). What is an inclusive city? Reconfiguring participation in planning with geospatial photovoice to unpack experiences of urban belonging among marginalised communities. *Visual Studies*, *39*(1-2), 144–163.

Cox, K. R. (1969). The voting decision in a spatial context. *Progress in Geography*, *1*, 83–117.

De Matteis, F., Bille, M., Griffero, T., & Jelić, A. (2019). Phenomenographies: describing the plurality of atmospheric worlds. *Ambiances*, *5*(5). https://doi.org/10.4000/ambiances.2526

Dewey, J. (1910). *How we think*. Digireads.com Publishing.

Dewey, J. (1929). *The Later Works of John Dewey, 1925 - 1953: 1929: The Quest for Certainty* (Vol. 4). Southern Illinois University Press.

Duncan, O. D., & Duncan, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review*, *20*(2), 210.

Esposito, E. (2022). *Artificial communication*. MIT Press.

Ess, C. (2017). Grounding internet research ethics 3.0: A view from (the) AoIR. In M. Zimmer & K. Kinda-Kurlanda (Eds.), *Internet research ethics for the social age.* . Peter Lang.

Gallup, G., & Rae, S. F. (1940). *The pulse of democracy: the public-opinion poll and how it works*. Simon & Schuster.

Girard, M., & Stark, D. (2007). Socio-technologies of assembly: Sense making and demonstration in rebuilding Lower Manhattan. In D. L. Viktor Mayer-Schönberger (Ed.), *Governance and Information Technology: From Electronic Government to Information Government* (pp. 145–176). MIT Press Cambridge, MA.

Halegoua, G. R. (2020). *The Digital City: Media and the Social Production of Place*. NYU Press.

Hartshorne, R. (1939). The nature of geography: A critical survey of current thought in the light of the past. *Annals of the Association of American Geographers. Association of American Geographers*, *29*(3), 173.

Jacomy, M. (2020). Epistemic clashes in network science: Mapping the tensions between idiographic and nomothetic subcultures. *Big Data & Society*, *7*(2), 205395172094957.

Joss, S., & Durant, J. (1995). *Public participation in science: The role of consensus conferences in Europe*. NMSI Trading Ltd.

Kant, I. (1787). *Kritik af den rene fornuft*. Det lille forlag.

Kitchin, R. (2023). *Digital Timescapes: Technology, Temporality and Society*. John Wiley & Sons.

Kitchin, R., Lauriault, T. P., & McArdle, G. (2015). Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards. *Regional Studies Regional Science*, *2*(1), 6–28.

Klinenberg, E. (2018). *Palaces for the people: How social infrastructure can help fight inequality, polarization, and the decline of civic life*. Crown.

Kornberger, M., Justesen, L., Madsen, A. K., & Mouritsen, J. (2015). *Making things valuable*. Oxford University Press.

Latour, B. (2005). Critical distance or critical proximity. *Unpublished manuscript.Available at Http://www.bruno-latour.fr/sites/default/files/P-113-HARAWAY.pdf.Accessed March*, *31*.

Lazarsfeld, P., & Fiske, M. (1938). The "panel" as a new tool for measuring opinion. *Public Opinion Quarterly*, *2*, 596–612.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). Big data. The parable of Google Flu: traps in big data analysis. *Science (New York, N.Y.)*, *343*(6176), 1203–1205.

Madsen, A. K. (2015). Tracing Data—Paying Attention. In M. et al Kornberegerm (Ed.), *Making Things Valuable*. Oxford University Press.

Madsen, A. K. (2018a). Dataficerede offentligheder: overvejelser over digitale metoder i lyset af den tidlige amerikanske pragmatisme. *Dansk Sociologi*, *29*(1), 97–119.

Madsen, A. K. (2018b). Data in the smart city: How incongruent frames challenge the transition from ideal to practice. *Big Data & Society*, *5*(2), 2053951718802321.

Madsen, A. K. (2022). Do you live in a bubble? Designing critical metrics to intervene in the cartography of urban diversity. *Projections*, *16*. https://scholar.google.com/citations?view_op=view_citation&hl=en&citation_for_view=53GI3fQAAAAJ:tS2w5q8j5-wC

Madsen, A. K. (2024). Digital methods as "experimental a priori" – how to navigate vague empirical situations as an operationalist pragmatist. *Convergence The International Journal of Research into New Media Technologies*, *30*(1), 94–115.

Madsen, A. K., Burgos-Thorsen, S., De Gaetano, C., Ehn, D., Groen, M., Niederer, S., Norsk, K., & Simonsen, T. (2023). The Urban Belonging Photo App: A toolkit for studying place attachments with digital and participatory methods. *Methodological Innovations*, *16*(3), 292–314.

Madsen, A. K., Grundtvig, A., & Thorsen, S. (2022). Soft City Sensing: A turn to computational humanities in data-driven urbanism. *Cities* , *126*. https://doi.org/10.1016/j.cities.2022.103671

Madsen, A. K., & Munk, A. K. (2019). Experiments with a data-public: Moving digital methods into critical proximity with political practice. *Big Data & Society*, *6*(1), 2053951718825357.

Madsen, A. K., Munk, A. K., & Søltoft, J. I. (2023). Friction by machine: How to slow down reasoning with computational methods. *Ethnographic Praxis in Industry Conference Proceedings*, *2023*(1), 65–68.

Madsen, A.K & Søltoft, J.I. (2023, December). *Interfacing with synthetic mundane politicians*. Generative Methods, Aalborg University.

Madsen, A.K, Søltoft J.I & Klitgaard, A. (2025). *Datafantasi: Fra styring til læring i en verden af vilde problemer*. In Nørlem, J., Klee, N. & Andersen, M.R.F (Ed.), *Organisatorisk læring og forandring*. Samfundslitteratur.

Marres, N. (2005). Issues spark a public into being: A key but often forgotten point of the Lippmann-Dewey debate. In B. L. A. Weibel (Ed.), *Making Things Public* (pp. 208–217). MIT Press Cambridge, MA.

Milgram, S., & Jodelet, D. (2014). Psychological maps of paris:(1970). In *The People, Place, and Space Reader* (pp. 45–49). Routledge.

Moats, D. (2021). Rethinking the "Great Divide." *Science & Technology Studies*, *34*(1), 19–42.

Moretti, F. (2013). *Distant reading*. Verso Books.

Munk, A. K., Meunier, A., & Venturini, T. (2019). Data sprints: A collaborative format in digital controversy mapping. In J. V. A. Ribes (Ed.), *digitalSTS: A Field Guide for Science & Technology Studies*. Princeton.

Munk, A. K., & Winthereik, B. R. (2022). Computational ethnography: A case of COVID-19's methodological consequences. In *The Palgrave Handbook of the Anthropology of Technology* (pp. 201–214). Springer Nature Singapore.

Nelson, L. (forthcoming 2026). Objectivity In/Through Partial Perspectives. In Madsen, A.K & Munk, A.K. (Ed.), *Handbook of digital and computational reserach methods in social science and humanities*. Edward Elgar Publishing.

Nelson, L. (2020). Computational grounded theory: A methodological framework. *Sociological Methods & Research*, *49*(1), 3–42.

Niederer, S., & Colombo, G. (2024). *Visual Methods for Digital Research: An Introduction*. John Wiley & Sons.

O'neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

O'Reilly, T. (2009). *What is web 2.0*. O'Reilly Media, Inc.

O'Reilly, T. (2011). Government as a Platform. *Innovations: Technology, Governance, Globalization*, *6*, 13–40.

Osborne, T., & Rose, N. (1999). Do the social sciences create phenomena?: the example of public opinion research. *The British Journal of Sociology*, *50*(3), 367–396.

Papazu, I., & Birkbak, A. (2022). *Democratic Situations*. Mattering Press.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin Press.

Pentland, A. (2010). *Honest signals: how they shape our world*. MIT press.

Perry, C. (2013). "The Neighborhood Unit." In *The Urban Design Reader* (pp. 78–89). Routledge.

Pilati, F., Munk, A., & Venturini, T. (2024). Generative AI for Social Research: Going Native with Artificial Intelligence. *Sociologica*, *18*(2), 1–8.

Pinel, C., & Svendsen, M. N. (2024). Domesticating data: Traveling and value-making in the data economy. *Social Studies of Science*, *54*(3), 429–450.

Rieder, B. (2013). *Problematic Facebook API permissions: the case of read_stream*. The Politics of Systems - Thoughts on Software, Power, and Digital Method. http://thepoliticsofsystems.net/2013/01/problematic-facebook-api-permissions-the-case-of-read_stream/

Shirky, C. (2009). *Here comes everybody: The power of organizing without organizations*. Penguin Books.

Søltoft, J. I., Kocksch, L., & Munk, A. K. (2024). Synthetic Interlocutors. Experiments with generative AI to prolong ethnographic encounters. In *arXiv*. arXiv. http://arxiv.org/abs/2410.11395

Surowiecki, J. (2005). *The wisdom of crowds*. Anchor Books.

Warner, M. (2021). *Publics and Counterpublics*. Zone Books.

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