

Trust for sale

A critical reading of the Norwegian National Strategy for AI

Kayla Holderbein

Vitenskapelig assistent, Institutt for medier og kommunikasjon, Universitetet i Oslo

kaylamh@uio.no

Steffen Krüger

Førstelektor, Institutt for medier og kommunikasjon, Universitetet i Oslo

steffen.kruger@media.uio.no

Christopher Wilson

Forsker, Institutt for medier og kommunikasjon, Universitetet i Oslo

christbw@uio.no

In January 2020, the Norwegian government published its first National Strategy for Artificial Intelligence (Norway, 2020). Norway was not the first to envision the possibilities of AI on a national level; recent years have seen an increasing number of countries publishing similar strategies. To date, at least 60 countries have published policy initiatives, according to the OECD (OECD, 2020). Norway, however, was the last Nordic country to publish such a strategy, following in the footsteps of Sweden, Denmark, and Finland who all made their strategies public in 2019 (Robinson, 2020).

Trust is an important theme in AI imaginaries and contemporary critiques of data-powered systems. Steedman et al. (2020) have noted a “data trust deficit” amongst the public, as the use and abuse of personal data becomes an ever-pressing topic of debate (818). These issues are also salient in the realm of AI, where data is increasingly brought to life as a very real part of the everyday, with algorithms operating in the background of most of our online interactions. Scholars such as Cath (2018) question whether we have the ethical and regulatory toolbox necessary to grapple with artificial intelligence—would (or, perhaps, *should*) human conceptions of trust suffice in the face of non-human mechanisms?

The theme of trust becomes particularly meaningful in the Nordic strategies. Trust is *the* distinct advantage that the Nordics claim to hold over other countries who have entered the AI race. Trust has been called “the Nordic Gold,” a unique factor that has led to the economic and social flourishing of these countries (Andreasson & Stende, 2019). Norway, in turn, is more commonly known for another kind of gold: *Oil* allowed the once-poor nation to grow into one of the wealthiest in the world. But as the age of oil dwindles, the nation needs to find equally profitable, but more sustainable industries. Data has been repeatedly called the ‘new oil’ in the Norwegian context (Bjellvåg, 2017), and Norway’s mass of public data in

combination with high levels of public trust give it “a key advantage in today’s global competition” (Norway, 2020, 6). Trust, transparency and openness are the very foundations upon which Nordic governments have been established, and these same cultural values have become a profitable resource in the global AI arena, as reflected in artificial intelligence strategies across Nordic nations (Robinson, 2020).

Focusing on one of these strategies—the Norwegian National Strategy for Artificial Intelligence (Norway, 2020)—this discussion paper seeks to spark debate around potential conflicts between public and private interests in artificial intelligence policy. It does so by addressing the main shortcomings and inconsistencies we find in the strategy regarding declared and demonstrated objectives, the positioning of public versus private interests, and assertions of AI neutrality. Each of these are introduced through excerpts from the strategy and considered in turn. We conclude by proposing opportunities to strengthen public AI policy.

Nordic-washing artificial intelligence

Norwegian society is characterized by trust and respect for fundamental values such as human rights and privacy. The Government wants Norway to lead the way in developing and using AI with respect for individual rights and freedoms. This can become a key advantage in today’s global competition (*Norway, 2020, 5–6*).

Values of trust, openness, and transparency have had a clear influence on the development of the Norwegian AI strategy; the word ‘trust’ alone is used 40 times throughout the document, and there is an entire sub-chapter devoted to ethics. This represents a strong emphasis on the virtuous handling of AI, indeed. However, a closer, more critical reading of the text suggests that this emphasis might be more lip-service than face-value. While the preface, introduction and the final parts of the strategy promote such concerns as human rights, ethics, the protection of individual privacy and freedoms as well as trustworthiness, these are practically absent from the document’s core sections. The effect is that of a smokescreen: beginning and ending with palatable public-facing ethical concerns while cutting to business matters in the document’s body. In this way, the ethical frame seems to function to soothe readers’ concerns, instead of providing substantial practical guidelines for AI implementation.

The strong evocation of trust and trustworthiness in the Norwegian strategy is more than a smokescreen, however. The document’s framing of trust effectively commodifies it: rendering it into the raw material that will fuel Norway’s success in the global AI industry. This is evident in the strategy’s presentation of Norway’s competitive advantage in the global AI market, which includes its “high level of public trust,” “a population and business sector that are digitally competent,” and “high-quality registry data that span over many decades” (Norway, 2020, 5).

If public trust and data are to be leveraged as capital, it is crucial that such business is conducted with constituents’ interests in mind. It is worth noting that the strategy’s audience is clearly stated as, “public and private entities seeking to develop and use artificial intelligence”—not civilians themselves (Norway, 2020, 2). This becomes ever-more salient in the core sections of the document, which guide potential industry partners through an understanding of AI’s opportunities and applications. The general public is not intended to be the document’s primary audience and is not addressed here. However, since its partici-

pation in the establishment of AI-driven services is of elementary importance to the success of such services, the public nevertheless plays a central role in the discourse on AI and society. This then begs the question of how, concretely and exactly, the public will be involved: What processes will be used to engage and educate? How will public values be protected and maintained in the face of business interests? And how will public trust be strengthened and justified (not simply harnessed) through AI initiatives?

Merely ticking the boxes of *trust* and *ethics* amounts to ‘Nordic-washing’: leveraging these values to further an agenda which ultimately may or may *not* align with the values themselves. As Cath (2020) emphasizes, *who* sets the agenda for AI governance and regulation is just as important as the values and logics embedded in such decisions (4). In this way, the strategy misses the opportunity to develop an integrated discourse that engages the public and acknowledges their role in its success.

Powering a new industry

The Government's goal is to facilitate sharing of data from the public sector so that business and industry, academia and civil society can use the data in new ways. Data can be regarded as a renewable resource (Norway, 2020, 13).

In an emerging global market such as that promised by AI, innovation is key for the growth and establishment of new industry. The hope is that “AI can lead to new, more effective business models and to effective, user-centric services in the public sector” (Norway, 2020, 5). As the strategy emphasizes, Norway hopes to facilitate innovation in AI through public-private partnerships and what the strategy terms, “digitalisation-friendly regulations” (Norway, 2020, 6). The first action suggested by the government in this regard, however, is to find opportunities for *de-regulation* of the business sector, removing obstacles that might, “hamper appropriate and desired use of artificial intelligence in the public and private sectors, so as not to hinder growth of the industry” (Norway, 2020, 6).

Once again, the initial presentation of this idea is more ‘public-friendly’ than later parts of the document imply. In our opinion, while some level of deregulation may indeed be necessary, this also presents challenges for accountability and transparency. AI involves the exercise of power, through the “discursive reinforcement of particular norms, approaches and modes of reasoning” (Beer, 2017, 11), and does so in a way that renders the power relationships behind AI invisible to the people that use it (see also Bucher, 2018; Cath et al, 2017). Therefore, instead of falling head-first into the notion of increased productivity through deregulation, we would encourage careful consideration of how regulatory measures strengthen and enable different social actors *differently*, and how this matters to the sustainability of AI policy in the long run. Along these lines, Steedman et al. (2020) call for, “collective, ecosystem solutions, for example, better regulation of data-driven systems, in order for them to be perceived as more trustworthy” (829). From their perspective, regulation of AI systems can act as a *trust-building* mechanism—which, in turn, could help strengthen the public’s perception and reception of national AI management. In that sense, smart regulation could be just as beneficial as deregulation, not only for industry interests, but for the public as well.

These sentiments are echoed in the UK House of Lords report *Digital technology and the resurrection of trust* (Select Committee on Democracy and Digital Technologies, 2020) which notes the public’s desire for greater governmental regulation of tech platforms and

data services in order to keep these parties *accountable* (34-35). According to the UK report, the challenge facing governments in holding data systems accountable is *keeping up* with the rapid pace of development (35). This is quite different from the Norwegian Strategy's concern about regulations hindering innovation. Rather, there is an opportunity to consider how accountability could improve public perception and engagement with AI systems. Grasping this opportunity would be very much in line with the Nordic precautionary principle of designing policy in order to avoid public harms (*føre-var-prinsippet*), as casually referenced in the introduction to the Norwegian Strategy (Norway, 2020, 2).

The “neutral” standard

Automation can also promote equal treatment, given that everyone who is in the same situation, according to the system criteria, is automatically treated equally. Automation enables consistent implementation of regulations and can prevent unequal practice (Norway, 2020, 27).

The use of data and algorithms to automate processes is foundational to artificial intelligence, and because this happens through algorithms beyond direct human control, AI is often thought of as *neutral* and *objective* (Bucher, 2018, 51). Yet, as Bucher notes, rule-based systems such as algorithms do not exist or act in a void; rather, they are created and interpreted by humans, for human purposes (52). The illusion of being ‘neutral’ or ‘beyond human control’ instead serves to conceal the human agency implicit in their functioning.

The Norwegian strategy acknowledges the limitations of rule-based algorithms and predictive technologies, highlighting the advantages of machine learning in contrast:

Unlike rule-based systems, where rules are defined by humans and are often based on expert experience, business logic or regulations, the concept of machine learning covers a range of different technologies where the rules are deduced from the data on which the system is trained. (Norway, 2020, 11)

It would be a mistake, however, to assert that machine learning avoids the problems of bias that accompany rule-based systems. As implied in the Strategy's discussion of data quality (57), human biases can be embedded in decisions about what datasets AI learns from, the kind of data that is collected, how data is valued or weighted, and how that data is used. In all these points, and others, human involvement is necessary (for a discussion, see Balaram, Greenham, & Leonard, 2018, 12).

The assertion that, “automation can also promote equal treatment” goes further, however, suggesting that neutrality is a precondition for *equality* (Norway, 2020, 27). Does neutrality imply equality? AI is only as good as the policies and norms it reinforces. It is a great amplifier of decisions already made, using past data to shape future realities (Bucher, 2018). This means that any *flawed* system criteria or regulations will also be repeated, until they are flagged for manual review.

This is not to say that automated processes should not be used; rather that neutrality may not always be beneficial. AI has the potential to be normative precisely *because* of its assumed neutrality— machine learning may continue to reinforce a set of norms or fall into discriminatory feedback loops, precisely because it lacks human subjectivity (Berscheid & Roewer-Despres, 2019; Rahwan, 2017). The role of a national strategy, we hold, should be to enable public sector agencies to consider these dilemmas in a way that can realistically approximate the ideals proposed above.

Conclusion: The ‘I’ in AI

Though its primary audience may be business and institutions, the Norwegian national strategy for AI does not ignore the citizen. On the contrary, the citizen, or more specifically, *citizen data*, is a tool for the strategy’s success. This is explicitly stated in the introduction: “Access to high-quality datasets is essential for exploiting the potential of AI” (Norway, 2020, 6). The use of all-inclusive terms like ‘datasets’ neutralizes and conceals where data come from—and data always come from people. In this instance, replacing the phrase “high-quality datasets” with “information about people” would dramatically shift the implicit connotations, but not the explicit meaning. Without the citizen, there would be no data—and Norway’s budding AI industry would fall flat.

As Balaram et al. (2018) emphasize, “When it comes to controversial uses of AI, the public’s views and, crucially, their values can help steer governance in the best interests of society. Citizen voice should be embedded in ethical AI” (10). Although citizens are vital to Norway’s AI success, the strategy seems to paint the individual as a passive contributor and beneficiary of AI—not an active participant. Of course, citizens are also the intended users of public AI technologies, and AI is intended to benefit them too. The intended uses are to strengthen the nation’s offerings within “health, seas and oceans, public administration, energy and mobility” (Norway, 2020, 7). So far, however, the benefits outlined in the strategy seem to be more heavily weighted toward profit and value production for industry and government. The result is a document not so much written *for* the individual as it is *selling* the individual as a means of profit. This begs questions about how the public will be included in the development and debate around artificial intelligence, and in what ways will the public be engaged in future policy? How will business and industry partners be encouraged to do the same?

To answer these questions, we might consider ways in which future AI policy and communications could be developed with the citizen in mind, and there are several resources to build on. In the UK, The Royal Society of Arts (RSA) has proposed a system of deliberative public dialogue, where the public is invited to ask questions, share input, and critique proposals (Balaram et al., 2018, 19). Rahwan (2017) has asserted a “society-in-the-loop (SITL)” system requiring, “various societal stakeholders to identify the fundamental rights that AI must respect, the ethical values that should guide the AI’s operation, the cost and benefit tradeoffs the AI can make between various stakeholder groups” (9). The participatory design community has suggested a number of methods for including societal values in AI infrastructures (Loi, Lodato, Wolf, Arar, & Blomberg, 2018). Norwegian scholars have proposed methods for including non-experts in technical discussions about the design of AI for policing (Vestby & Vestby, 2019). These approaches incorporate subjective, human engagement in AI processes, and represent the best mechanisms for ensuring that AI systems manifest the values of trust, transparency and openness espoused by the Nordic Strategies for AI (Robinson, 2020). They deserve as much attention as the Norwegian Strategy pays to deregulation and global competition.

References

- Andreasson, U., & Stende, T. (2019). Nordic municipalities’ work with artificial intelligence. In *Nordic municipalities’ work with artificial intelligence*. <https://doi.org/10.6027/no2019-062>
- Balaram, B., Greenham, T., & Leonard, J. (2018). *Engaging citizens in the ethical use of AI for automated decision-making*. Retrieved from https://www.thersa.org/globalassets/pdfs/reports/rsa_artificial-intelligence---real-public-engagement.pdf

- Beer, D. (2017). The social power of algorithms. *Information Communication and Society*, 20(1), 1–13. <https://doi.org/10.1080/1369118X.2016.1216147>
- Berscheid, J., & Roewer-Despres, F. (2019). Beyond transparency. *AI Matters*, 5(2), 13–22. <https://doi.org/10.1145/3340470.3340476>
- Bjellvåg, S. (2017). Data er den nye oljen. Retrieved February 5, 2021, from Dagens Næringsliv website: <https://www.dn.no/teknologi/sommerkjappe/microsoft-norge/kimberly-lein-mathisen/-data-er-den-nye-oljen/2-1-125484>
- Bucher, T. (2018). *If... then?: algorithmic power and politics*. Oxford: Oxford University Press.
- Cath, C. (2018). Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180080. <https://doi.org/10.1098/rsta.2018.0080>
- Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2017). Artificial Intelligence and the ‘Good Society’: the US, EU, and UK approach. *Science and Engineering Ethics*, 24(2), 505–528. <https://doi.org/10.1007/s11948-017-9901-7>
- Loi, D., Lodato, T., Wolf, C. T., Arar, R., & Blomberg, J. (2018). PD manifesto for AI futures. *ACM International Conference Proceeding Series*, 2, 18–21. <https://doi.org/10.1145/3210604.3210614>
- Norway. (2020). *Strategi Nasjonal strategi for kunstig intelligens*. <https://www.regjeringen.no/no/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/>
- OECD. (2020). The OECD Artificial Intelligence Policy Observatory . Retrieved February 5, 2021, from OECD.AI website: <https://oecd.ai/>
- Rahwan, I. (2017). Society-in-the-Loop: Programming the Algorithmic Social Contract. *Ethics and Information Technology*, 20(1), 5–14. <https://doi.org/10.1007/s10676-017-9430-8>
- Robinson, S. C. (2020). Trust, transparency, and openness: How inclusion of cultural values shapes Nordic national public policy strategies for artificial intelligence (AI). *Technology in Society, preprint*, 101421. <https://doi.org/10.1016/j.techsoc.2020.101421>
- Select Committee on Democracy and Digital Technologies. *Digital Technology and the Resurrection of Trust*. , (2020).
- Steedman, R., Kennedy, H., & Jones, R. (2020). Complex ecologies of trust in data practices and data-driven systems. *Information Communication and Society*, 4462. <https://doi.org/10.1080/1369118X.2020.1748090>
- Vestby, A., & Vestby, J. (2019). Machine Learning and the Police: Asking the Right Questions. *Policing: A Journal of Policy and Practice*, 0(0), 1–15. <https://doi.org/10.1093/police/paz035>