Trust in Public Sector AI: The Case of Finland's Aurora National Programme

Author(s): Dan, Sorin; Kalliokoski, Johanna; Shahzad, Khuram

Title: Trust in Public Sector AI: The Case of Finland's Aurora National Programme

Year: 2024

Version: Accepted manuscript

Please cite the original version:

Dan, S., Kalliokoski, J. & Shahzad, K. (2024). Trust in Public Sector AI: The Case of Finland's Aurora National Programme. In J. Paliszkiewicz & J. Gołuchowski (Eds.), *Trust and Artificial Intelligence: Development and Application of AI Technology* (pp. 242-257). Routledge studies in trust research. Routledge. https://doi.org/10.1108/S2398-391420240000007003

Chapter 20: Trust in Public Sector AI: The Case of Finland's AuroraAI National Programme

Sorin Dan
School of Management, University of Vaasa, Vaasa, Finland
sorin.dan@uwasa.fi, ORCID: https://orcid.org/0000-0003-0565-7282

Johanna Kalliokoski
School of Management, University of Vaasa, Vaasa, Finland
johanna.kalliokoski@uwasa.fi, ORCID: https://orcid.org/0000-0002-3959-2758

Khuram Shahzad
School of Technology and Innovations and Innovation and Entrepreneurship InnoLab, University of Vaasa, Vaasa, Finland
khuram.shahzad@uwasa.fi, ORCID: https://orcid.org/0000-0002-6452-0879

Citation:
Dan, S., Kalliokoski, J. and Shahzad, K. (2024) Trust in Public Sector AI: The Case of Finland's Aurora National Programme, in Paliszkiewicz, J. and Gołuchowski, J. (eds.) *Trust and Artificial Intelligence Development and Application of AI Technology*, **Routledge:** London, **Jufo 2**, https://www.routledge.com/Trust-and-Artificial-Intelligence-Development-and-Application-of-AI-Technology/Paliszkiewicz-Goluchowski/p/book/9781032626321.

Abstract

This chapter explores the relationship between trust and public-sector AI. It examines the implications for trust of an AI program, AuroraAI, established by the Government of Finland that aimed to improve citizen centricity of public services. The chapter reviews the program's possible public benefits, risks, and ethical dilemmas and finds that the design of citizen-centric services with the help of AI-enhanced data analytics is likely to result in cost reduction, better service experience, and improved effectiveness. Yet, these benefits are accompanied by public risks and ethical dilemmas that are characteristic of AI and require further examination.

Keywords: Trust, public sector, artificial intelligence, Finland

Introduction

Governments at national, regional, and local levels have taken steps to experiment with and implement artificial intelligence (hereafter AI) technologies to improve the organization and provision of government services across different sectors and policy areas (Misuraca & van Noordt, 2020). AI is part of a growing trend in public management scholarship and practice, which see AI as one technology that is part of a set of 'highly integrative and immersive technology' (Kim et al., 2022, p. 362) that are contributing to a deepening of digital government. Smart technology, i.e. cloud computing, big data analytics, social media, the Internet of Things (IoT), and AI, may operate according to government as a platform (GaaP) principles (O'Reilly, 2011; Brown et al., 2017) that value leaner interfaces, user-centricity, and a different role for the government which becomes 'a convener and an enabler rather than the first mover of civic action' (O'Reilly, 2011, p. 15). Such an approach to public management claims to enhance public trust through the use of inter-connected digital infrastructure that creates new and rich information flows, leading to increased accessibility of government information and services (Cordella & Paletti, 2019).

AI is assumed to lead to a redesign of service delivery processes towards a renewed and more trustworthy model that places the service user at the centre (O'Reilly, 2011). Increased citizens' demand for efficient and personalized public services has surged public sector investments in new AI-based digital technologies. Therefore, how AI is designed and implemented in practice in the complex world of public sector organizations, which oftentimes lack citizens' trust, becomes essential to achieving this aspiration. Majorly, citizens are unaware of how the AI system makes decisions that affect them, leading to mistrust (Wang et al., 2021). To foster trust in government organizations, it necessitates ensuring that AI-driven public services can satisfy the citizen's expectations and ensure transparency and accountability.

Yet, despite significant interest in AI from academics and policymakers alike, there is insufficient research on the implications of using AI for trust in governmental organizations that operate in a sensitive political environment and are under public scrutiny (Vogl et al., 2020; Zuiderwijk et al., 2021). Although some studies (e.g. Wirtz & Müller, 2019) have attempted to point out the influence of AI on public value creation theoretically, there remains a significant lack of an empirical framework examining AI implementation in the public sector to foster citizen trust.

This chapter addresses the implications for the trust of experimenting with AI in public sector organizations by exploring the National Artificial Intelligence Programme AuroraAI (hereafter AuroraAI) developed by the Government of Finland. This program ran from January 2020 until December 2022 (Ministry of Finance, 2023).

Trust and public sector AI

The research on trust and public sector AI builds on the extensive literature on trust that is found in the field of public administration and management (Nyhan, 2000; Van de Walle et al., 2008). This research is also influenced by the digital government research on citizen and user centricity in digital and smart services and government as a platform (Lee & Lee, 2014; Pope, 2019; Rana et al., 2013). How to design and provide public services in a way that responds to citizens' needs and adds value to them is an important theme within this literature. Digital government scholars have argued that Web 2.0 has created the conditions for adopting a platform thinking in public governance that better integrates service provision across different providers, yet to this day its application has largely remained government-centric instead of citizen-centric (Anttiroiko et al., 2014; Sigwejo & Pather, 2016).

The use of ubiquity principles and intelligent solutions may address this challenge and better personalize the design and delivery of public services (de Sousa et al., 2019). Big data and data analytics reinforced by AI techniques may facilitate the creation of new public service models that place citizens at the centre and that incorporate a greater emphasis on prevention and responsiveness (Wirtz & Müller, 2019). There is some empirical evidence that points in this direction, yet the existing evidence is limited and contradictory. Zheng et al. (2018) find that the use of AI can lead to service integration, improved civil service satisfaction, and, importantly, reduced waiting times for citizens and users. Automation of service delivery processes can free up case workers' time which can be used to provide more personalized service in high-risk cases where it is most needed. Coupled with the adoption of self-service delivery through chatbots, this service model can enhance the availability and centricity of public services (Eggers et al., 2017). AI-enhanced data analytics can provide predictive information across policy areas and types of services that can be used to develop better preventive healthcare, educational and employment services (Fernandes et al., 2018; Sun & Medaglia, 2019). Yet, the extent to which AI-supported techniques can result in greater citizen centricity and trust depends on the nature and complexity of service, considering that AI is best suited for managing caseloads and administrative procedures that are labor intensive, time-consuming and which require the processing of a large volume of data (Kleinberg et al., 2018).

Furthermore, AI holds great promise for effective government services, improved decision-making processes, and optimizing resource allocation to interact with citizens and deliver services (Wang et al., 2021). Existing studies (e.g. Wirtz & Müller, 2019; Sun & Medaglia, 2019) have highlighted the potential positive influence of AI on public value creation, as the goal of governmental organizations is to improve citizen-government interaction by developing citizens' trust for their delivered services. Trust in public sector AI, often referred to as public trust, is considered a key indicator of public value and explains the level of confidence citizens have in the government's procedures for AI-based services (Wang et al., 2021). Citizens tend to have faith in AI-based services' ability to address their concerns while acting in their best interests. This leads to citizen acceptance of such digital services by the government which enhances the public tendency to interact with government organizations (Hu et al., 2021), as well as trust. Several scholars (e.g. Bertot et al., 2010) have already highlighted the importance of information and communication technologies (ICTs) in fostering public trust, as such technologies offer citizens timely information which is a key factor to develop citizen's perception of transparency and trust in government (Wang et al., 2021). Prior research has also highlighted some AI implementation challenges in the public sector, such as AI algorithms have the potential to unintentionally create bias, perpetuate historical inequalities, exhibit a preference for specific political perspectives, or strengthen unwanted behaviors (Janssen & Kuk, 2016), leading to creating uncertainty among citizens. While Wang et al. (2021) found no significant relationship between the effective use of AI and trust in government, suggesting the inconsistencies in value orientation between governments (intended purpose to promote IT) and citizens' behavior (i.e. perception of meeting the needs).

Thus, despite significant interest in AI from academics and policymakers alike, there is insufficient research on the implications of using AI for trust in governmental organizations that operate in a sensitive political environment and are under public scrutiny (Vogl et al., 2020; Zuiderwijk et al., 2021). Existing research majorly remained scattered and somewhat limited with more theoretical analysis of AI and public trust while also highlighting the related challenges of AI implementation, leading to uncertainty. Therefore, it is important to address such limitations in existing research as the implications of AI to develop trust in governmental organizations is a critical endeavor that corroborates the responsible use of AI technologies in safeguarding citizens' rights and interests.

Methodology

This chapter employs a qualitative single case study research design. As the chapter addresses a real-life phenomenon, a case study design is beneficial for exploring its implications as applicable to the novel setting of public service delivery (Eisenhardt, 1989). Data collection involved semi-structured interviews with key individuals involved in program design, ideation, and implementation, and the analysis of relevant policy documents and existing evaluations. The interviewees represent different professional backgrounds, such as ethics experts, digital analysts, and other technology professionals in municipal governments and ministries (see Appendix).

All interviewees were chosen on purpose and invited to participate in the study due to their involvement and/or knowledge in the case study. The interviews were held online during March-May 2020. Selected follow-up interviewing was conducted in February 2021 to assess further developments in implementation in municipal government that took place between May 2020 and February 2021. All interviews were recorded and transcribed verbatim, consisting of 42 pages of transcribed material. In addition, we performed an analysis of relevant documents produced by different actors, such as government, consultancies, and academic institutions.

Finland's AuroraAI and the quest for trust in public sector AI

The Finnish context is conducive to AI experimentation. As a Nordic welfare state, local governments and well-being services counties are legally mandated to provide a considerable array of services in a relatively sparsely populated geographic area. Finland boasts a high degree of digital literacy and innovation and has been ranked first for two years in a row in the European Commission's Digital Economy and Society Index (European Commission, 2020). However, the aging of the population coupled with fiscal pressures on state finances, have determined policymakers to utilize the potential of AI for public service improvement. This interest is also motivated by the realization that the current public service model is not only fiscally unsustainable but also ineffective at using large amounts of existing data for analytical and predictive purposes to improve policymaking and citizen outcomes (AuroraAI Development and Implementation Plan, 2019).

To address these challenges and opportunities, over the past few years, the Finnish authorities turned to AI to examine its possible societal role. This led to the publication of several policy documents such as the 2017 Research and Innovation Council's report which proposed an ambitious vision and a roadmap according to which Finland aspires to become 'the most attractive and competent environment for experimentation and innovation' by 2030 (Research and Innovation Council Finland, 2017, p. 1). This vision emphasizes the role of data and AI as key drivers for building a platform economy and boosting digital innovation in public services. As a result, Finland became one of the first countries to launch a national AI programme, entitled Finland's Artificial Intelligence Programme. During 2017–2019, the programme published a series of reports including *Finland's Age of Artificial Intelligence*, which assessed the importance and likely impact of AI and proposed recommendations to help the nation succeed in the age of AI (Ministry of Economic Affairs and Employment, 2017).

The AI developments surveyed above culminated with the launch of a dedicated public sector AI program, AuroraAI, which was a national, multi-annual program that is officially defined as 'a concept for human-centric and ethical society in the age of artificial intelligence' (AuroraAI

Development and Implementation Plan, 2019, p. 9). The ambition of AuroraAI is to transform Finland into a ‘human-centric society’. It sought to achieve this by means of a new service model that is based on a ‘human-centric service market’ that gravitates around particular life events and business activity (ibid., p. 12). It uses an artificial neural network (ANN), ‘the AuroraAI network’, which is defined as a ‘decentralized open network of smart services and applications’ that connects different service providers (public, private and non-governmental) (ibid., p. 9). Its citizen-centric focus and service orientation is clear: ‘AuroraAI seeks to provide a holistic set of personalized AI-driven services for citizens and businesses in a way that is human-centric and works towards their wellbeing as its ultimate goal’ (OECD AI Policy Observatory, 2021).

Data is constructed around people’s service needs, user segments, and service history logs, which enables the AI system to suggest personalized and targeted services seamlessly. It aims to facilitate both an improved user experience and more effective services that respond to changing and complex service requirements and preferences. This seamless service experience means that it is possible to use AuroraAI in multiple ways such as in the form of a website, email, chatbot conversation, or other messaging tool, much like in the case of the internet (AuroraAI Development and Implementation Plan, 2019). Instead of having one single AI system, it is suggested that AuroraAI would form a network of smart services that can better adapt to technological developments (Cybercom Finland Oy, 2019). Through the management of their personal data, service users are supposed to gain more holistic access to available services and better ‘situational awareness’ of their multi-faceted well-being. The program defined situational awareness as ‘a data-based model of describing statuses, attributes, trends, and system dynamics relating to customers or the analyzed phenomenon in sufficiently realistic detail to enable its use in customer/phenomenon-focused management’ (AuroraAI Development and Implementation

Plan, 2019, p. 9). Through 'situational management' users can access tailored public services based on their service chain history instead of them 'figuring out' what services they need or without a public authority suggesting such service.

Public benefits, public risks, and ethical dilemmas

We present our exploratory results of the implications of AuroraAI for trust by grouping them into three categories: public benefits, public risks, and ethical dilemmas.

Public benefits

The first expected public benefit of AuroraAI is cost reduction in public services. By better customizing service provision, it is possible to reduce costs over the medium and long term through the optimization of service offerings. This involves delivering the right service at the right time and to the right person in a citizen-centric way. Within AuroraAI, cost reduction, however, is not a stand-alone objective; rather it is closely connected to the expectation that effectiveness will improve if services proactively address citizens' needs. Using data analytics supported by AI, service providers can harness data about citizens' holistic well-being and provide service recommendations that can improve well-being, as this interviewee explains:

'The effect of current public service structures on society is not strong enough and that needs to be boosted and the only way to boost it is to go human-centric, which in this particular context means creating all these snapshots of states of well-being in certain population segments and quite possibly with individuals even, and then building these service ecosystems around those segments to provide services specifically targeted, and preferably proactive services, to address people's needs maybe even before they realize they have those needs.'

Based on personalized data, it is possible to identify risk factors in advance design tailored interventions, and address wellbeing. In this service model cost reductions may result from the fact that it costs less to prevent an unfavorable life situation using AI than to respond to it once it has occurred (Sun & Medaglia, 2019). Happier and healthier people will also use the service less, which may bring about savings over the long run. In addition, efficient optimization of service provision and available resources could reduce spoilage and oversupply (AuroraAI Development and Implementation Plan, 2019). AuroraAI's service chatbots and automated user-experience interface may reduce the working hours of public service employees, who could use the saved time to manage higher-risk cases. These changes have the potential to improve citizen-centricity and service effectiveness to the extent that case workers' time is dedicated to people who need it the most and benefit most from more personalized care.

Another public benefit concerns improved user experience. If service provision happens mostly digitally then it means that it would be more flexibly available than today: AuroraAI could be accessed anytime and from anywhere. This can reduce the need for other means of communication which are time-consuming for both service providers and users. In addition, it can also make the process faster for the user: instead of waiting in line on the phone or spending time searching on the internet, AuroraAI can provide guidance almost immediately. While the current suomi.fi service brings together information from several public services to a single website, the AuroraAI experience would be different for the user due to an automated customization process. Moreover, user experience is more seamless as AuroraAI integrates service offerings from different providers to create a holistic service plan and address, in a coordinated way, a particular life event. In this renewed service model, the user doesn't need to be aware of all the services or procedures that are available to choose the most suitable one, which simplifies user experience:

'If the system succeeds… then of course it will make public services qualitatively better. It will be a fluent and flexible way to combine all kinds of sources of information and if the system is developed in the right way, it can indeed help you to fix your problems in a better way than it has been originally possible.'

As an AI public service program, AuroraAI can add significant public value by using large amounts of user data that currently many different Finnish public service organizations are legally mandated to collect. However, this data is not integrated across different service areas and public organizations and is insufficiently used in policy making. This deprives policymakers and service providers of useful knowledge that may significantly improve policy effectiveness, as this interviewee notes:

'As you have noticed, I'm… a bit cynical about how public administration has used data so far. There's much room for improvement there… For example, social [and] healthcare sector organizations have enormous amounts of data, but we don't really from, my point of view, at the moment have smart ways to understand the data and to use the data. (…) If we have, for example, data about some issue over ten or fifteen years… we can use that to also predict what's going to happen in the future.'

However, existing GDPR and national legislation currently do not allow user-level data sharing across different organizations. To address this legal challenge, the Finnish law of automated decision-making in public services is currently under examination (Vainio et al., 2020). If these are resolved, AuroraAI could constitute a more effective way to utilize the data and use it to predict changes in population segments. In addition, the AuroraAI system gathers data itself in the form of history logs to customize service provision according to service needs identified in population segments:

'Perhaps it also gives feedback to the public sector authorities in a way that has not been possible before because all the records about uses can be analyzed later and used for correcting the system. It can use the feedback information to develop the system better and better in the long run.'

Public risks

AuroraAI is a system that lives on data, which includes both personal and individual data and data derived from population clusters. Without this data, the system cannot function and its success depends on the accuracy and quality of data that it receives and generates. This gives rise to several public risks and ethical dilemmas. First, there is a risk that the estimated cost reduction is not as significant as predicted. If AuroraAI does not acquire quality data to target high-risk cases, then it would serve those individuals who are already doing well and who may be more willing to share their personal data given that data sharing is voluntary. In Finland most social and healthcare sector costs are used for serving a small part of the population (Valtiontalouden tarkastusvirasto, 2017). This implies that without effective targeting of people in need, cost reduction is unlikely to be significant. Yet, this proactive targeting may be difficult to realize, and it depends on the extent to which people in need make use of AuroraAI's service model. This public risk challenges the very notions of human centricity and societal inclusion and raises the question of how to design AI systems in a way that they are appropriate and effective for the most vulnerable members of society.

Another significant public risk lies in the clustering of the data and the creation of user segments, a process that can be disempowering in itself (Dan & Kalliokoski, 2023). Algorithms in AI applications may contain bias due to errors and may result in unfavorable outcomes for some users (Panch et al., 2019). In addition, despite the existence of big data (which may not be accurate enough), it is difficult to confirm that individuals accurately match the user segments with which they are compared. This is particularly problematic in the case of AuroraAI which aims for a holistic concept that incorporates various facets of well-being (economic, social, emotional, spiritual, etc.). Moreover, it raises the question of how being a 'person at risk' or otherwise shapes the user experience of the system:

'The relationship between segment and individual has never been clearly defined…, and can I deny that I'm part of some segment? Can I say no, I'm not part of that segment? You're misreading me, or even if I am statistically part of a segment and the data categorizes me into a segment it's some segment that I don't want to be a part of… at the moment.'

A significant risk is posed by legal bottlenecks concerning the full implementation of AuroraAI. In the context of Finland, this risk is particularly relevant considering that similar hurdles led in 2019 to the resignation of the government following the crash of a major social and healthcare reform (Hiilamo, 2020). Relatedly, policymakers' quest for creating citizen-centric services by means of AI capabilities is equally challenging considering that AuroraAI's service network consists of providers and organizations (public, private, and non-governmental) that have different organizational values, norms, and operating procedures. In this case, it can be challenging to provide holistic support to the individual while sharing responsibilities and assessing performance in a multi-actor network:

'The Ministry of Finance wants to build a human-centric society in which AI is used for the good of the individuals. But all the organizations that are in the network have their own responsibilities. There are laws that say that okay Kela [Finland's Social Insurance Institution], that's your role. Vero [Finnish Tax Agency], that's your role. City of Tampere. You have your responsibilities. How do you turn this into a human-centric society?'

The multi-actor system also raises public risks related to privacy and data management. AuroraAI is assessed as mostly secure for the user by the interviewees, yet data security depends on how each participating organization handles user data. For the system to assess an individual's complex situation accurately and to intervene proactively, it needs relevant, holistic data from different sources. This poses questions concerning user consent and control of the data, especially given that the data may 'speak against the user', which can undermine the incentive to provide reliable data. AuroraAI architects claim that all personal data that citizens provide to the system is anonymized and forms a digital twin that is not identical to the data provider (AuroraAI Development and Implementation Plan, 2019). However, previous research has shown that anonymity may be undermined when combining data from multiple sources, which is particularly relevant for a sparsely populated country such as Finland (Honkanen, 2020; Green et al., 2017). In addition, interviewees noted that it might be tempting for the state to use the system for surveillance purposes, given its proactive stance that targets behavioral change (Dan & Kalliokoski, 2023):

'One big concern that is all the time existing is how to prevent the formation of a big brother society.'

Ethical dilemmas

Ethical dilemmas involve decision-making between at least two ethical imperatives, none of which is preferable or acceptable (Yung, 2016). Our empirical results point to several ethical dilemmas.

A significant ethical dilemma concerns the personal data conundrum and its implications for effective citizen-centric services. While it is possible to use AuroraAI without sharing personal data, users cannot use its full citizen-centric functionalities unless they share at least some personal data that is relevant and accurate. There are trade-offs between the type and amount of personal data that is shared and the accuracy of service recommendations that AuroraAI delivers in response. 'Anonymous browsing' is possible within AuroraAI, according to the interviewees, but results may lack precision and personalization, which undermine the very concept of the program. Moreover, to ensure a degree of anonymity (which it can be argued it cannot be absolute), AuroraAI needs to securely have in place technical requirements such as blocking IP addresses and phone IDs and refraining from using identifying cookies. Individuals need to consent to the processing of their data to use AuroraAI and they are responsible for what they share with the system, yet they need to understand its functionalities to decide what data they share about themselves and for what purposes. The dilemma concerns the extent to which the service providers who make up the AuroraAI network should be legally mandated to use population data for issuing service recommendations given possible public benefits and risks that may be different from one individual to another.

'The thing is that the [artificial] neural network alone will not solve anything. It will never make things better than they are now. It will need good quality data and it will also need... a certain amount of that data. Not necessarily open data [which means] data from all the organizations that is shared with everyone, but the [artificial] neural network cannot function without quality data.'

This implies that for AuroraAI to be effective, particularly in the case of vulnerable citizens, Finnish people have to trust and accept the service model and its data management principles and procedures. This brings us to the next dilemma, which is that some users might have difficulties understanding the AuroraAI concept and to whom they give their personal information and consent given its multi-actor logic. AuroraAI can become a black box, unintelligible not only to its designers and architects but particularly to the most vulnerable members of society who may need it the most, including children, immigrants, and persons with disabilities (see also Haataja &

Latvanen, 2019). It should be noted that even though Finland is a digitally advanced nation and in general possesses above-average digital skills, not all citizens have the same level of digital literacy to make use of AuroraAI capabilities.

Another salient ethical dilemma that results from the interviews lies with the predictive features of AuroraAI. By using data analytics to analyze social and health datasets, it is possible to predict risk factors for specific population segments based on historical data. Although the Finnish legislation does not allow this use of data, the ethical dilemma concerns whether this should be allowed, given its possible public benefits and risks. Such predictions can have an impact on how people live their lives. Even if the aim is to improve well-being for the common good, an ethical dilemma remains concerning whether AI should be affecting people's behavior in the first place and whether people have the right not to know the predicted trend if they so wish.

Although the usage of AuroraAI is not planned to be mandatory, nor does any individual have in most cases any responsibility to follow the suggestions created by the system, it must be acknowledged that AuroraAI is most likely going to be trusted by people who use it, and it may well have an impact on people and their decision making. If created and used properly, an AI system may be less biased and contain more information than, for example, in the case of a human counselor (Kleinberg et al., 2019), but an ethical dilemma is apparent concerning the system's power to 'voice' recommendations and the implications that this new reliance on an AI system has over the medium and long term. Will it only assist with personalized service recommendations or will it go further and further, like a train going on a one-way track, to decisively influence and determine lives and destinies?

'(...) In Aurora there's this application that they're creating now in which young people [students] around the age of 14 – 15 are advised that you should apply to this school or you should go to

study this and then the question that raises is… how much the state actually controls the future of those kids (….). It is a system that categorizes those kids into A type and B type and so on. That's the danger (…).'

There is an inherent danger that the use of AI and data analytics becomes so essential to policymaking, and perhaps the only approach to policymaking, that governments may wish to overstep their boundaries. This concern was well reflected in the interviews:

'It will be so tempting for the public sector to cut costs and do all kinds of things to make the sector more efficient by using the data in a way that makes it as a surveillance society and whether or not I give my permission for them to do so doesn't make the surveillance any less of a surveillance.'

If citizens consent to the use of data, which in a high-trust Nordic welfare state such as Finland, is likely to be the case, then where does the limit go? A regulatory framework and democratic processes that follow transparency and accountability principles can provide checks and balances (Floridi et al., 2018; Wirtz & Müller, 2019), yet public governance arrangements may become so entrenched in AI that they become oblivious to correcting their own faults. In so doing they may form what elsewhere we call *'surveillance public governance'* (Dan & Kalliokoski, 2023), a system in which power relations between the state, networked service providers, and service users are significantly shaped by the use of AI, to their own benefit and risk.

Conclusions

This chapter has explored the relationship between public sector AI and trust by analyzing the AuroraAI National Programme implemented by the Government of Finland. Our main findings show that the creation of AI-supported services is likely to lead to public benefits in terms of systemic cost reduction and improved user experience and service effectiveness. This results from the delivery of tailored service recommendations and offerings through the AuroraAI network according to service needs for particular life events and population segments. Cost reductions and improved effectiveness may result from the use of predictive capabilities within AuroraAI based on which policymakers and service providers can issue service recommendations ahead of time that prevents the occurrence of unfavorable and costly future life events.

Nevertheless, our results show that the attainment of these benefits is fraught with public risks and ethical dilemmas that are inherent in AI technologies. Public risks that are apparent in the context of AuroraAI include the unequal use of the system across population groups with those who need it the most being at risk of using it the least, and thus benefitting from it, the least. This is particularly relevant given the data intensity of the program, which requires accurate, personalized high-quality data to be effective. Without the provision, sharing, and availability of such data, the very concept of citizen-centric public services is at risk. An additional significant public risk relates to the clustering of the data and creation of population segments, which can be both disempowering and difficult to apply considering that inclusion in population clusters can be an imprecise process that can result in error. AuroraAI has raised ethical dilemmas related to the government's use of AI, which despite its aim of improving the common good, can lead to public surveillance at the expense of individual liberty (Dan & Kalliokoski, 2023).

References

Anttiroiko, A. V., Valkama, P., & Bailey, S. J. (2014). Smart cities in the new service economy: building platforms for smart services. *AI & society*, 29, 323–334.

AuroraAI Development and Implementation Plan (2019). *AuroraAI – Towards a human-centric society. Development and implementation plan 2019–2023 based on the preliminary study on the Aurora national artificial intelligence programme*. Ministry of Finance, Helsinki, Finland.

Bertot, J. C., Jaeger, P. T., & Grimes, J. M. (2010). Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly*, 27(3), 264–271.

Brown, A., Fishenden, J., Thompson, M., & Venters, W. (2017). Appraising the impact and role of platform models and Government as a Platform (GaaP) in UK Government public service reform: Towards a Platform Assessment Framework (PAF). *Government Information Quarterly*, 34(2), 167–182.

Cordella, A. & Paletti, A. (2019). Government as a platform, orchestration, and public value creation: The Italian case. *Government Information Quarterly*, 36(4), 1–15.

Cybercom Finland Oy (2019). *Arkkitehtuurikuvaus AuroraAI: Työpaketti 4* [Architectural description of AuroraAI: Work package 4]. Report.

Dan, S. & Kalliokoski, J. (2023). "Surveillance Public Governance": Towards a new public sector reform trajectory? Working paper, School of Management, University of Vaasa, Vaasa, Finland.

de Sousa, W. G., de Melo, E. R. P., Bermejo, P. H. D. S., Farias, R. A. S., & Gomes, A. O. (2019). How and Where Is Artificial Intelligence in the Public Sector Going? A Literature Review and Research Agenda. *Government Information Quarterly,* 36(4), 1–14.

Eggers, W. D., Fishman, T., & P. Kishani (2017). *AI-augmented human services: Using cognitive technologies to transform program delivery*. Deloitte Center for Government Insights.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, 14(4), 532–550.

European Commission (2020). *Digital Economy and Society Index 2020*. Brussels: European Commission.

Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & G. Van Erven (2018). Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. *Journal of Business Research,* 94, 335–343.

Floridi, L., Cowls, J., Beltrametti, M. et al. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds & Machines* 28, 689–707. https://doi.org/10.1007/s11023-018-9482-5

Green, B., Cunningham, G., Ekblaw, A., Kominers, P., Linzer, A., & Crawford., S. (2017). *Open Data Privacy*. Berkman Klein Center for Internet & Society Research Publication. http://nrs.harvard.edu/urn-3:HUL.InstRepos:30340010.

Haataja, M. & Latvanen, M. (2019). *AuroraAI-esiselvityshanke - Etiikka-työkokonaisuuden suositukset* [AuroraAI preliminary study – Recommendations of the Ethics-work package]. Report.

Hiilamo, H. (2020). Why did Social and Healthcare Services Reform Fail in Finland? *Socialmedicinsk Tidskrift*, 97(3), 433–441.

Honkanen, M. (2020). *Omadata osana AuroraAI-hanketta* [MyData as a part of AuroraAI-project]. Report.

Hu, Q., Lu, Y., Pan, Z., Gong, Y., & Yang, Z. (2021). Can AI artifacts influence human cognition? The effects of artificial autonomy in intelligent personal assistants. *International Journal of Information Management*, 56, 102250.

Janssen, M., & Kuk, G. (2016). The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly*, 33(3), 371–377.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *The Quarterly Journal of Economics*, 133(1), 237–293.

Kim, S., Andersen, K.N., & Lee, J. (2022). Platform Government in the Era of Smart Technology. *Public Administration Review*, 82, 362–368.

Lee, J. & Lee, H. (2014). Developing and validating a citizen-centric typology for smart city services. *Government Information Quarterly,* 31(1), 93–105.

Ministry of Economic Affairs and Employment (2017). *Finland's age of artificial intelligence - Turning Finland into a leading country in the application of artificial intelligence. Objective and recommendations for measures.* Publications of Ministry of Economic Affairs and Employment, MEAE reports 41/2017. Helsinki, Finland.

Ministry of Finance (2023). *National Artificial Intelligence Programme AuroraAI*. Retrieved 13.9.2023. https://vm.fi/en/national-artificial-intelligence-programme-auroraai.

Vainio, N., Tarkka, V., & Jaatinen, T. (2020). *Assessment memorandum on the need to regulate automated decision-making within public administration in general legislation*. Publications of the Ministry of Justice, Reports and guidelines 2020:14. Helsinki, Finland.

Misuraca, G. & van Noordt, C. (2020). *Overview of the use and impact of AI in public services in the EU*. Publications Office of the European Union: Luxembourg.

Nyhan, R. C. (2000). Changing the Paradigm: Trust and its Role in Public Sector Organizations. *The American Review of Public Administration*, 30(1), 87–109.

OECD AI Policy Observatory (2021). *AuroraAI*. Retrieved 15.2.2021. https://oecd.ai/dashboards/policy-initiatives/2019-data-policyInitiatives-24206.

O'Reilly, T. (2011). Government as a Platform. *Innovations: technology, governance, globalization*, 6(1), 13–40.

Panch, T., Mattie, H., & Atun, R. (2019). Artificial intelligence and algorithmic bias: implications for health systems. *Journal of global health*, 9(2), 020318.

Pope, R. (2019). *Playbook: Government as a platform*. Ash Center for Democratic Governance and Innovation, Harvard Kennedy School.

Rana, N.P., Dwivedi, Y.K., & M. D. Williams (2013). Evaluating alternative theoretical models for examining citizen centric adoption of e-government. *Transforming Government: People, Process and Policy*, 7(1), 27–49.

Research and Innovation Council Finland (2017). *Vision and road map of the Research and Innovation Council Finland*.

Sigwejo, A. & S. Pather (2016). A Citizen-Centric Framework for Assessing E-government Effectiveness. *EJISDC*, 74(1), 1–27.

Sun, T. Q. & Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly*, 36(2), 368–383.

Valtiontalouden tarkastusvirasto (2017). *Tuloksellisuustarkastuskertomus - Paljon palveluja tarvitsevat ja käyttävät asiakkaat perusterveydenhuollossa* [Performance audit report - Frequent users of basic health care services]. Valtiontalouden tarkastusviraston tarkastuskertomukset 11/2017. Helsinki, Finland.

Van de Walle, S., Van Roosbroek, S., & Bouckaert, G. (2008). Trust in the public sector: is there any evidence for a long-term decline? *International Review of Administrative Sciences*, 74(1), 47–64.

Vogl, T. M., Seidelin, C., Ganesh, B., & Bright, J. (2020). Smart Technology and the Emergence of Algorithmic Bureaucracy: Artificial Intelligence in UK Local Authorities. *Public Administration Review*, 80(6), 946–961.

Yung, B. (2016). Introduction: Ethical Dilemmas, Social Values, and Public Policy: The Context of Governance and Citizenship. In: Yung B. & Kam Por Y. (eds.): *Ethical Dilemmas in Public Policy. Governance and Citizenship in Asia*, pp. 1–19, Springer, Singapore.

Wang, C., Teo, T. S., & Janssen, M. (2021). Public and private value creation using artificial intelligence: An empirical study of AI voice robot users in Chinese public sector. *International Journal of Information Management*, 61, 102401.

Wirtz, B.W. & Müller, W.M. (2019). An integrated artificial intelligence framework for public management. *Public Management Review,* 21(7), 1076–1100. doi:10.1080/14719037.2018.1549268.

Zheng, Y., Yu, H., Cui, L., Miao, C., Leung, C., & Yang, Q. (2018). SmartHS: An AI platform for improving government service provision. Paper presented at the 32nd AAAI Conference on Artificial Intelligence, 7704–7711.

Zuiderwijk, A., Chen, Y-C., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government Information Quarterly,* (28)3, 101577.

Appendix: Key interviews

INTERVIEW	*TIMING*	*MALE (M)/FEMALE (F)*	*AREA OF EXPERTISE*
1	March 2020	M	Local governance (LG)
2	May 2020	M	Digitalisation (D)
3	May 2020	M	Technology, ethics (TE)
4	May 2020	F	Science, ethics (SE)
5	February 2021	M	Local governance (LG)