Devising Regulatory Sandboxes and Responsible Practices for Designing AI-based Services in the Finnish Public Sector

Nitin Sawhney[0000-0002-8088-914X] and Ana Paula Gonzalez Torres[0000-0002-2793-7501]

Aalto University, Department of Computer Science,
Konemiehentie 2, 02150 Espoo, Finland

nitin.sawhney@aalto.fi, ana.gonzaleztorres@aalto.fi

Abstract. In this paper, we examine the opportunities and challenges of devising mechanisms for responsible AI practices in the public sector. We discuss how innovations in public sector AI must comply with suitable ethical and regulatory frameworks, while fostering experimentation and participation of diverse actors throughout the *AI lifecycle* of designing, deploying, and assessing such services. We examine AI-based innovations and ethical practices in the context of prototype systems and end-user services provided by two Finnish public sector organizations. The paper proposes *regulatory AI sandboxes* engaging multiple stakeholders and integrative frameworks (*MLOps* and *RegOps)*, to facilitate experimentation, co-learning, and deployment of future public sector AI services with ethical practices and regulatory compliance. Upcoming regulations such as the AI Act and the AI Liability Directive make the adoption of such sandboxes beneficial as a space for limited experimentation, that goes beyond ethical considerations to specific legal compliance and allocation of liability. In the future, this framework must be validated through pilot experimentation with public/private stakeholders and regulators in different areas of high-risk AI services.

Keywords: regulatory sandboxes, public sector, AI-based services, ethical practices, regulatory compliance

1 Introduction

The public sector is increasingly embracing algorithmic decision-making and data-centric infrastructures to improve digital services in areas such as education, healthcare, and urban mobility (Floridi, 2020; Ruckenstein *et al.*, 2020; Haataja *et al.*, 2020). As such *Public AI Services* become more prevalent and affect people's lived experiences, the inevitable socio-political, ethical, and legal implications require a more critical understanding of the rights, risks, and responsibilities for both the providers and recipients of such services (Crawford, 2021; Sawhney, 2022). Much of the current discussion around the use of big data, computational systems, and artificial intelligence in urban contexts centers around so called "smart cities" and intelligent infrastructures (Aurigi & DeCindio, 2008; Foth, 2018; Hollands, 2008; Kitchin, 2019). Despite the benign intentions of public and private actors many crucial ethical implications must be examined (Kitchin, 2016). Some AI-based systems are being used by governments for biometric surveillance, criminal justice, and other forms of citizen monitoring, which pose higher risks for abuse and unfair incrimination (AI Now, 2018; AI Watch, 2020) if they are not made easily transparent, accountable, or their legitimate use challenged by civil society. Engaging multiple stakeholders and civil society in participatory deliberations on the use of AI has begun to be explored in recent pilot projects (Rolfe, 2019; RSA, 2019; Sawhney, 2020).

The Finnish public sector is undergoing such an algorithmic transformation, with several AI-based services already in use such as book recommendation services in public libraries and chatbots advising maternity and child health, while new services are being planned for many high-risk domains such as tax administration and immigration services (Ruckenstein et al., 2020). Recent government reports have claimed that Finland would integrate AI systems into virtually every part of society, from healthcare to industry in an open and ethical way (AI Finland, 2019). The City of Helsinki is taking a concerted ethical stance in introducing new data-centric and algorithmic services; in 2020 it launched the *AI Register* (Haataja *et al.*, 2020) which aims to transparently inform citizens about algorithmic services provided by the city (Floridi, 2020). In July 2021, the City of Helsinki collaborated with the Berkman-Klein Center's Research Clinic (BKC, 2021), inviting international researchers to develop new approaches that emphasize citizen participation and human oversight when introducing public AI services. Innovative AI-based services in the Finnish public sector must not only take into account ethical considerations to ensure they are fair, unbiased, trustworthy, and accountable, they must also be designed in compliance with relevant municipal, national, and EU-based regulatory policies and frameworks.

Finnish digital infrastructures and policies are usually harmonized with regulatory guidance from the European Union. On April 21, 2021, the European Commission published a proposal for regulating artificial intelligence systems developed, placed and used in the EU market, called the Artificial Intelligence Act (EC, 2021). While there is a vigorous debate around the efficacy and practical feasibility of the proposal, the consultations and deliberations with EU member states are continuing to refine and implement this framework as EU-wide regulations in the future. The diverse and contested discourses around the AI Act among legal/policy experts (OECD, 2021), regulators, as well as private (DIGITALEUROPE, 2021) and public actors, offer a timely window of opportunity to critically examine the challenges and concerns emerging, while promoting multi-stakeholder engagement, citizen participation, and civic agency in shaping the AI Act and its compliance framework both in Finland and the EU. Should the proposal become law, governments and municipalities must assess and adapt their algorithmic services to comply with the new regulations, which call for stricter measures of transparency, accessibility, and accountability.

In this paper, we examine the opportunities and challenges of devising mechanisms for responsible AI practices in the public sector. We discuss how innovations in public sector AI must comply with suitable ethical and regulatory frameworks, while fostering experimentation and participation of diverse actors throughout the *AI lifecycle* of designing, deploying, and assessing such services. We examine AI-based innovations and ethical practices in the context of prototype systems and end-user services provided by two Finnish public sector organizations. The paper proposes *regulatory AI sandboxes* engaging multiple stakeholders and integrative frameworks (*MLOps* and *RegOps)*, to facilitate experimentation, co-learning, and deployment of future public sector AI services with ethical practices and regulatory compliance. In the future, this framework must be validated through pilot experimentation with public/private stakeholders and regulators in different areas of high-risk AI services.

2 Adoption of AI-based Systems in the Public Sector: Challenges and Opportunities

The public sector plays a unique role in the socio-cultural, economic, and political context of society that distinctly underpins its mandates, values, practices, governance, relationships, and trust with other actors (citizens/recipients, private sector, non-governmental, and state actors). Hence, the ways in which algorithmic and data centric approaches, increasingly constituted into AI-based systems and services, are designed, adopted, used, and assessed by the public sector vs. private enterprises often require different considerations, challenges, and opportunities. A key driver for the public sector is the *rule of law* that dictates how it achieves public good for its diverse stakeholders. "In both roles of user and producer of AI-based solutions, the public sector's choices are determined by specific policies and it operates within and in compliance with the given legal mandates provided by the rule of law (e.g. understanding scrutiny and accountabilities, apply equity, transparency, consistency in decisions and redress whenever needed)" (Manzoni *et al.*, 2022). Hence, the adoption of new technologies and how they are used by the public sector are directly influenced by distinct legal provisions, administrative roles, and ethical responsibilities to serve key functions that affect all citizens and non-citizens, including vulnerable and marginalized communities. While profit, market dominance, and shareholder value dictate the thrust of private enterprises, the public sector must operate within different mandates of legal, regulatory, and ethical values that include equity, inclusion, transparency, accountability, and good governance. However, to achieve its aims the public sector relies on collaboration and consultation with a complex ecosystem of private sector companies, non-governmental organizations (NGOs), as well as state and civil society actors in the local, national, and international sphere, to effectively deliver services and perform its activities. The public sector must also exercise its power for functions, granted to them through legislation, in conformity with *public administration law*, as well as ethical and human rights principles (Leslie *et al.*, 2021). Hence, the benefits, impacts, and risks associated with introducing novel digital technologies and services, in particular those leveraging big data and AI-based systems, must be carefully assessed, integrated, and rolled-out by the public sector to align with the governance frameworks, legal principles, regulatory compliance, and ethical values it is deemed responsible for (OECD, 1999).

Innovative AI-based public services can aim to support a range of context-specific public values encompassing *operational public values* e.g., efficiency and effectiveness, *political public values* e.g., citizen participation, equity, and accountability, and *social public values* e.g., inclusion, trust, and sustainability (Barker *et al.*, 2021;

Manzoni *et al*., 2022). Digitization and AI-based systems in the public sector can potentially support not only more effective delivery of public services, but also foster greater trust and interaction with its recipients and stakeholders, improved decision-making processes for better policy outcomes, and to help optimize internal management of key functions (Manzoni *et al*., 2022). Conversational AI systems using Natural Language Processing (NLP) or multimodal voice interaction can support open-ended inquiries and improve accessibility for users with visual impairments or limited digital literacy. Such AI-based chatbots can also collect user opinions and enhance citizen participation in defining public sector services (Androutsopoulou *et al*., 2019). Urban mobility data pooled from diverse sources (historical and real-time data) can contribute to better urban planning and modelling; AI algorithms can learn dominant patterns of movement in a city at different times of day, predict demand, and support participatory design of suitable transportation alternatives and pathways for urban mobility with diverse stakeholders in the city (Abduljabbar, 2019; Sawhney, 2022).

Yet with these assumed benefits come many potential risks associated with algorithmic bias, surveillance, infringement of people's privacy and rights, lack of inclusion of all sectors of society (due to inadequate digital access or literacy), inequitable treatment or lack of fairness in how services may be rendered, and inadequate transparency, accountability, and redress for erroneous outcomes. Hence, the public sector tends to take a more conservative, slow, and cautious approach in incorporating AI-based technologies into their systems or introducing them widely for societal use, not to mention the challenges imposed by their lack of internal capacity or expertise in AI, the scaling and robustness of systems needed to effectively serve societal functions, and the necessary technical validation, legal compliance, and multi-stakeholder consultation often required. These conditions offer distinct challenges and opportunities for the design of novel platforms, methods, practices, and ethical/regulatory approaches to support public sector AI innovation.

The challenges of adopting and consequently deploying AI-based solutions require engaging responsible and ethical practices with multiple stakeholders involved across the entire AI lifecycle (De Silva & Alahkoon, 2021). The literature highlights the need to further research agile and dynamic tools for technical and responsible adoption across this lifecycle, for instance use of MLOps in the public sector (Pechtor & Basl, 2022). This is a particular challenge as frameworks that guide AI projects in the public sector need to take into consideration changing political priorities and legislative constraints, insufficient internal expertise in IT and AI development, and the lack of common data standards and appropriately curated data resources (Mehr, 2017). In terms of data governance, one of the main barriers to the uptake of AI projects in the public sector is the complex regulatory environment they must operate in. For instance, when developing any big data or algorithmic service the public sector must take into consideration, at the very least, the 'General Data Protection Regulation' (GDPR) where "large volumes of data from diverse sources require complex negotiated agreements between different stakeholders on how to control which data are collected, how they are collected, in which format they are stored, and who has access to it" (Harrison & Luna-Reyes, 2020; Medaglia *et al.*, 2021). In the near future, the public sector must also contend with the upcoming legal directives such as 'Data Act', 'Data Governance

Act', 'AI Act' and 'AI Liability Directive', as they pertain to their datafication and AI-based services.

AI-based systems in the public sector are often developed across a multi-stakeholder ecosystem with some aspects devised in-house, procured as software or services from the private sector, or the result of public-private partnerships for co-development (Engler & Renda, 2022), each of which impose different sets of requirements and obligations under the regulatory measures that pertain to each actor. The public sector is typically based on an organizational logic of hierarchy and verticality (Pūraitė *et al.*, 2020), while working within an AI lifecycle benefits from horizontal embedding of roles and responsible actions from multiple stakeholders across its different stages. In high-risk domains (such as the financial banking sector), regulatory sandboxes have been used to explore the possibilities and implications of algorithmic systems before wider deployment, allowing for piloting, monitoring and experimentation in a highly controlled environment in conjunction with multiple-stakeholders and regulatory experts, thereby limiting risks on a larger scale (Manzoni *et al*., 2022). In later sections of this paper, we examine how these *regulatory AI sandboxes* can be devised in conjunction with MLOps software frameworks, responsible practices, and ethical principles to facilitate innovative public sector AI services, particularly in the Finnish context.

3 Transitioning from Ethical Practices to Regulatory Compliance in the Public Sector

The necessity for critically examining ethical issues in public sector AI has arisen due to recent high-profile events that have sparked public outrage regarding their harmful impacts. For instance, the use of biased algorithms in criminal risk assessment in the U.S. (ProPublica, 2026), the use of algorithms that downgraded student examination grades in the U.K. (Algorithm Watch, 2020), or the use of 'discriminatory algorithms' by the Dutch Tax Authority for the provision of childcare benefits (Netherlands Court of Audit, 2022). A study on the global landscape of AI ethics guidelines showed a general convergence towards values of transparency, fairness, non-maleficence, responsibility, privacy, beneficence, autonomy, trust, and sustainability (Jobin *et al*., 2019). The European Commission set up a High-Level Expert Group on Artificial Intelligence to develop 'Ethics Guidelines for Trustworthy AI', which focused on seven key requirements: human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental wellbeing, and accountability (HLEGAI, 2019).

While ethical principles have been regarded as the first step, a crucial concern is how to transform them into regulatory measures. This has been raised in parallel with an awareness that ethical AI should be embedded throughout the entire AI lifecycle (European Parliamentary Research Service, 2020). In regulatory terms, the proposed AI Act has seen the arduous task of implementing ethical AI by means of requirements and obligations within a structure that builds a regulatory compliance scheme in terms of the AI lifecycle. Thus, the transition from ethical practices to regulatory compliance is to be embedded in the six stages of design, employment of data, development, deployment, maintenance, and retirement (De Silva & Alahkoon, 2021). Nevertheless, at least for the public sector, the transition from ethical principles to regulatory compliance must take into consideration the responsibility that institutions have towards citizens and, as such, incorporate a responsible AI approach that combines strict legal compliance with ethical AI, as the later encompasses regulations that support societal needs (European Parliamentary Research Service, 2020).

The complexity of the public sector also comes from the involvement of different stakeholders throughout the AI lifecycle. As such, while public sector organizations could be involved in the design of AI-based services, a third party may supply data for AI training (European Parliament, 2020), while development may be contracted to another private sector entity that provides the AI software infrastructure. Thus, to embed ethical AI practices in the process each stakeholder involved in the process would bear certain roles and responsibilities in each stage of the AI lifecycle. For instance, in the design stage what if any AI-based approaches should be adopted for a particular context (Dignum, 2022; Kela, 2019); in the training of datasets, ensuring that the data is unbiased (Doshi-Velez & Kortz, 2017; Kela, 2019; European Parliament, 2020); in the development stage whether there is appropriate verification, testing, and validation of outcomes (Stirbu, Granlund & Mikkonen, 2022); in the deployment stage what harmful or unforeseen impacts may emerge (Demos Helsinki, 2022); during maintenance, whether the use of an AI-based system creates discriminatory feedback loops (D'Ignazio, 2020); and finally during retirement, considering implications of recalling a system that people have come to rely on. Such ethical considerations must also be embedded in future implementations to address the constraints of regulatory compliance imposed by the proposed AI Act for high-risk AI systems. These include training, testing, and evaluation of AI systems with quality data (article 10), human oversight to prevent or minimize risks (article 14), adoption of mechanisms to address unintended feedback loops (article 15), constant monitoring through the lifecycle (article 9, 10, 12, 17 and 29), and withdrawal or recall of non-compliant high-risk AI systems (article 21).

4 Examining AI-based Principles, Practices & Strategies in the Finnish Public Sector

Finland is known to excel in providing digital public services, devoting more than half of its digital budget to the digitalisation of public services (DESI, 2022). The Finnish context is distinct in the close attention paid not only to capitalize on AI-capabilities but also to support a responsible human-centric approach (Ministry of Economic Affairs and Employment, 2021). For instance, under the Prime Minister's Office the Finnish Government commissioned a study that provides an assessment framework and policy recommendations to support public governance in their adoption of non-discriminatory AI and algorithmic decision-making (Demos Helsinki, 2022). Thus, the Finnish AI approach in the public sector can offer insights on how to reconcile the needs for innovative public services with responsible practices that embeds legal and ethical considerations throughout the AI lifecycle.

We have examined AI-based projects at Kela, the Social Insurance Institution of Finland, that provides social security benefits for Finnish residents. KelaLab, a part of Kela's Innovation Unit, has experimented with speech analytics, classification of documents and customer messages using machine learning (ML). We invited KelaLab to present their work at the 'Data, AI, and Public Sector Governance' conference held on September 15, 2022 to examine a case study of their experiments in using ML to identify customer needs in messages received on their website. Their implementation of ML was intended to address the challenge of managing vast volumes of messages received, given scarce human resources and to ensure urgent messages are handled in a timely manner. ML classification allowed messages to be sent to appropriate channels based on categories inferred such as 'in need of advice', 'give information' or 'urgent' (Mattila & Juuti, 2022). The benefits of using AI-based system improved speed and consistency, while it posed challenges like lack of explainability as it was not clear what sentence the model based its decisions on and the subjectivity of labeling messages according to the classifications categories.

Kela has also devised principles on ethical use of AI for benefit handling, developed to guide the development of AI-based systems according to 'transparency', 'human-centered approach', and 'responsibility' (Kela, 2019). In line with such principles KelaLabs has also taken into consideration whether the use of ML techniques can result in the exclusion of people whose messages do not fit the classification labels of input training data and whether such implementations can introduce implicit biases (Mattila & Juuti, 2022). The case study shows that ethical principles are ingrained in the experimental innovation initiatives but do not generally go beyond considerations at an abstract level. While such efforts are commendable it raises the issue of how the public sector can comply with concrete provisions of emerging regulations and how ethical principles can be embedded throughout the AI lifecycle in a multi-stakeholder ecosystem. For instance, Kela's principles were developed in 2019 before the EC released its proposed AI Act, which has specific requirements and obligations that surpass general ethical principles. Thus, there is a need for the public sector to expand its approach to developing responsible AI-based systems.

A second example is the City of Helsinki, which under their 'Data Strategy' intends to leverage advanced analytics, such as ML, dynamic optimization, and predictive models to improve their operations and the use of public resources. As a result, their strategy has been data-driven while their flagship innovation has been the deployment of chatbots for various public services. So far, the city has deployed at least seven different chatbots to provide citizens with information related to sports facilities and services, parking related questions, book recommendations, healthcare support, maternity guidance, and rental apartment search. Even though these chatbots are rule-based and do not currently employ AI approaches, they provide an example of how the city has proactively implemented their framework for the responsible utilization of AI systems based on a 'metadata model'. The model was designed to collect different elements regarding the operation of algorithms used in order to make the decision logic behind the services understandable to the 'city's customers'. It provides a concrete example on how to translate ethical principles to regulatory compliance in the AI Act, which requires 'transparency and provision of information to users' (article 13).

Moreover, a recent partnership between the City of Helsinki and Berkman Klein Center at Harvard University developed a model for ethical AI support in municipal schools, with models based on participatory design and human accountability (Berkman Klein Center, 2021). In this case, the use of AI by the City of Helsinki aimed to 'improve personalized learning programs, student well-being, and retention within Helsinki's vocational education and training programs'. The City of Helsinki established ethical guidelines to 'prohibit the categorization of students by metadata'. As part of the study's outcomes, it was highlighted that policymakers that are considering the integration of AI should take a participatory approach and foster trust, especially if intended for sensitive areas (e.g., schools, housing, transportation, law enforcement). Finally, the efforts to develop AI policy for the City of Helsinki led to considerations of cooperation across municipal structures, accountability, and human oversight through the AI lifecycle. Thus, like Kela, the City of Helsinki has incorporated high-level considerations into their future innovation strategies, but it will have to translate them into concrete compliance measures as required by the proposed AI Act.

The key challenge that arises in the public sector is that there is a constant need to demonstrate innovation over short periods of time (Pechtor & Basl, 2022) and that there is a lack of connection between public administration bodies as they work in silos (Pūraitė et al., 2020). For instance, a lack of communication between units experimenting with AI-based innovation and those that deploy them into production can lead to a disconnect in how such systems should be effectively rolled-out, validated and maintained in practice beyond a pilot context (Stirbu, Granlund & Mikkonen, 2022). Experimentation allows public administrations to understand citizens' reactions, user behavior and gather feedback for their prototype concepts and pilots in a controlled manner (Schaefer, Lemmer & Kret, 2021). It highlights the need for tools and platforms that facilitate experimental AI-based systems that are both technologically innovative, ethically responsible, and legally compliant.

5 Devising Regulatory AI Sandboxes for Experimentation in the Public Sector

Experimentation is needed for the critical exploration of both technological and legal implications of AI systems with diverse stakeholders before such projects can be transitioned into wider deployment in society. For instance, under the proposed AI Act, public sector use of AI-based solutions would be categorized as high-risk and thus, to be lawful, would have to comply with a requirement for 'data and data governance' which refers to the need for quality data for 'training, validation, and testing data sets' (art. 10). This requirement then relates to the previously mentioned current and upcoming data regulations. Developing an understanding of the inter-relations between the different regulatory frameworks, at the European level for the regulations and the national level when it comes to the legal directives, would be better served by an environment that allows for supervised implementations instead of after-the-fact correction of unwarranted faults in compliance by means of administrative fines (art. 71).

In this regard, *regulatory sandboxes* can serve as controlled environments regulated by law, which can be virtual and/or physical and allow validation and verification of AI-based solutions, exploring both technical and regulatory compliance (Manzoni *et al*., 2022). Such a controlled environment is typically under the supervision of a regulator for a limited period to allow technical learning from development and testing of innovations in a 'real-world environment', as well as legal experimentation with regulatory regimes to understand the limits of regulatory requirements and obligations (European Parliament, 2022). Since 2021, regulatory sandboxes have begun to be adopted to tackle the regulation of new technological advancements due to agile and flexible nature for experimentation. The proposed AI Act in title V establishes sandboxes as 'measures in support of innovation'. Article 53(1) devises AI regulatory sandboxes as "a controlled environment that facilitates the development, testing and validation of innovative AI systems for a limited time before their placement on the market or putting into service pursuant to a specific plan." In such sandboxes the direct supervision and guidance by the competent authorities offers a safe harbor to ensure compliance with regulatory requirements and relevant legislation supervised.

Thus, sandboxes address the need for experimentation in terms of compliance with new regulations within a framework of controlled risks and supervision, while seeking to improve regulators' own understanding of new and emerging technologies. However, the EC Proposal in article 52(4) indicates that participants in AI regulatory sandboxes would remain liable for any harm inflicted on third parties as a result of experimentation in the sandbox environment; this imposes the same liability regime for both sandboxed and non-sandboxed high-risk AI applications. In this regard, an environment that allows for all stakeholders to interact and contribute to the design and development of an AI system, with an ability for transparent monitoring and auditing can facilitate the tracing and allocation of potential liability concerns that can be flagged or assigned during experimentation. This can implicitly reduce the risk of harms (to end users) and liability (to providers) by limiting their scope and impact, while allowing all parties including regulators to understand the scenarios and limitations of AI technologies and the relevant regulatory measures.

Such sandboxes when specifically devised for experimentation with AI-based systems and services, in conjunction with continuous monitoring via software frameworks, leads to the potential for deploying *regulatory AI sandboxes* in the public sector. Devising AI-based innovations and responsible practices can be better facilitated through an approach that engages multiple stakeholders throughout the lifecycle of design, development, validation, and regulatory compliance; this allows for a more comprehensive approach that ensures a thorough understanding on part of all actors involved that is not just focused on 'rapid prototyping' (Pechtor & Basl, 2022), but instead on responsible long-term AI innovation with higher impact. This requires a more dynamic means of integrating the AI development lifecycle and ethical design practices using contemporary software frameworks such as *Machine Learning Operations (MLOps)*, which support continuous monitoring, versioning, enhanced transparency, auditing, and improved usability of resulting AI systems (Ranawana & Kuranananda, 2021).

MLOps capitalize on creating automated pipelines that can allow one to 'continuously integrate, train, and deploy new versions of models quickly, effectively, and seamlessly without any manual intervention'; such ML pipelines automatically realize functions like 'data ingestion, data transformation, continuous ML model (re-)training, (re-)deployment, and out presentation' (Ruf *et al.*, 2021). This can be extremely useful for environments with constantly changing needs in the context of regulatory sandboxes, as they are established for a short period of time with the aim of experimenting with different viable paths towards regulatory compliance. One of the benefits of MLOps is they can reduce technical efforts expended on certain aspects of an ML pipeline, like data management tasks, by means of automation. Such automation could pose an obstacle when trying to modify the deployed models to adjust for requests from regulatory bodies like future AI European Board or Member States competent authorities. In such regards, aspects of MLOps like continuous integration and continuous deployment can help add new features that comply with regulatory body requests to a deployed model more rapidly. According to the Finnish technology company Solita the use of MLOps has been increasing since 2018 but entities still lack best practices for using it in the context of responsible AI.

Regulatory Operations or RegOps have emerged for use in high-risk domains, such as medical devices, as specialized software frameworks for creating 'an environment where the regulatory burden is carried out by the automated process' (Granlund & Vedenpää, 2022). Recent work by Solita has explored integrating MLOps and RegOps pipelines to support continuous design, training, and regulatory processes as a concept prototype for AI-based medical device software development (Stirbu, Granlund & Mikkonen, 2022). The prototype system was designed to be deployed in a production environment with its machine learning (ML) models trained during development but placed in a "locked" state so that its outcomes cannot be changed dynamically for regulatory reasons. With proper regulatory oversight in a controlled sandbox environment, such ML models could be retrained more dynamically to improve outcomes, with suitable versioning, auditing, verification, and compliance validation at each stage as needed. This would support piloting of new features, models, and datasets while monitoring its technical and regulatory compliance, while allowing both developers and regulators to continuously learn from iterative experimentation.

To support dynamic collaborative means for regulatory compliance, Saidot, a Finnish software firm, has developed a platform for AI governance and transparency while developing best practices as 'policy templates' (Saidot, 2022). Through their platform and practices Saidot seeks to support a comprehensive approach for key actors to engage in the AI lifecycle. They plan to develop governance tools that monitor every stage of an AI project to facilitate regulatory compliance from design to retirement. According to Saidot, the challenge of setting up dynamic models for regulatory compliance is the integration needed with an organization's machine learning models and AI systems to aggregate relevant data at different stages; this would allow for devising suitable triggers that require auditing, verification, validation or deliberations around regulatory compliance. For instance, continuous monitoring could allow detection of data drift that then triggers the event "the organization has to do X and Y to comply with regulatory requirements". In each domain of use one must carefully define what sort of events trigger regulatory compliance measures. In a scenario where there are multiple providers and developers, the dynamic monitoring of an AI system through its lifecycle by means of RegOps would better facilitate the tracing of harms and liability for systems outputs causally related to system attributes or behaviors created by individual stakeholders. Experimentation using AI sandboxes with dynamic lifecycle monitoring can hence target the intent, causation, and mitigation of liability in high-risk scenarios without stifling technology innovation (Truby *et al.* 2022).

As previously seen, the space for adoption of AI-based solutions in the public sector has several complications. From an ecosystem of multiple stakeholders involved throughout the lifecycle of an AI project, and complex regulatory environment, to specific needs for dynamic technical and compliance tools. Considering the need for guided and controlled experimentation under regulatory sandboxes, it must be noted that the AI Act has a requirement for 'human oversight' (art. 14). Such a requirement can provide the necessary connection for the implementation of sandboxes that allow experimentation with different levels of automation and human involvement either in terms of MLOps for the technical needs of AI-based solutions and RegOps as a tool for dynamic compliance. Thus, by means of sandboxes the public sector can benefit from more proactive exploration with innovative AI-based systems or services that would otherwise not be attempted because of their categorization as 'high-risk'. Through the means of regulatory AI sandboxes, the various public and private sector actors involved in the AI development lifecycle, whether they provide the necessary infrastructure for MLOps, offer regulatory compliance tools like RegOps, deploy and support the systems, or validate their compliance, can carefully experiment with and examine the technological and legal implications of using AI based services in the public sector under both existing or forthcoming complex regulatory frameworks.

6 Conclusions

The public sector faces distinct challenges to incorporate AI-based systems for improving its services that include not only developing suitable technological capacity and expertise, tools and datasets needed, but also devising responsible AI practices, ethical principles and means for compliance with the complex regulatory environment it faces. There are opportunities to develop partnerships with private sector enterprises and engage multiple stakeholders in devising innovative AI-based services, however they also demand suitable frameworks for design, development, and validation to ensure they align with ethical principles such as transparency, equity, inclusion, accountability, and good governance. Furthermore, because of upcoming regulations such as the AI Act and the AI Liability Directive we believe that regulatory AI sandboxes provide a space for limited experimentation that goes beyond ethical considerations to specific legal compliance and allocation of liability. For experimental and high-risk domains regulatory AI sandboxes, that combine MLOps and RegOps frameworks with responsible practices, offer a viable approach that supports continuous experimentation and learning across the AI lifecycle in conjunction with multiple stakeholders including developers and regulators for both technical validation and regulatory compliance. This offers a suitable pathway from experimental prototyping to production deployment that may foster responsible innovations in public sector AI.

References

AI Finland. (2019). Leading the way into the era of artificial intelligence: Final report of Finland's Artificial Intelligence Programme 2019, *Ministry of Economic Affairs and Employment* 2019:41.

AI Now Institute (2018). AI Now Report 2018, Tech. Rep., [Online] www.ainowinstitute.org.

Algorithm Watch (2020), Automating Society Report 2020, [Online] https://algorithmwatch.org/en/automating-society-2020/.

Annex III, High-Risk AI Systems referred to in Article 6(2), in Annexes to the Proposal for a Regulation of the European Parliament and of the Council. Brussels, 21.4.2021, COM(2021) 206 final.

Aurigi, A & De Cindio, F. (eds). (2008). Augmented urban spaces: articulating the physical and electronic city. *Ashgate Publishing.*

Berkman Klein Center for Internet & Society, Harvard University. (2021). Open Access Resources for AI in Schools. [online] https://bit.ly/3oeIeM7.

City of Helsinki, 2020. [Online] https://ai.hel.fi/en/ai-register/

Crawford, K. (2021). Atlas of AI. *Yale University Press.*

De Silva, D. & Alahakoon, D. (2022). An artificial intelligence life cycle: From conception to production. *Patterns,* 10 6, 3(6), pp. 1-13.

Demos Helsinki (2022). Promoting equality in the use of Artificial Intelligence - an assessment framework for non-discriminatory AI, *Policy Brief 2022:25*, http://www.tietokayttoon.fi/en.

DIGITALEUROPE (2021). Initial findings on the proposed AI Act. [online] https://bit.ly/3zRKQ51.

Dignum, V. (2022), Responsible Artificial Intelligence – from Principles to Practice. [Online] https://arxiv.org/abs/2205.10785.

Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S. J., O'Brien, D., Scott, K., Shieber, S., Waldo, J., Weinberger, D., Weller, A. & Wood, A. (2017) Accountability of AI Under the Law: The Role of Explanation, *Berkman Center Research Publication,* http://dx.doi.org/10.2139/ssrn.3064761.

Engler, A. C. & Renda, A. (2022). Reconciling the AI Value Chain with the EU's Artificial Intelligence Act. *CEPS In-Depth Analysis.*

European Commission (2022), Digital Economy and Society Index (DESI) 2022, Finland, https://digital-strategy.ec.europa.eu/en/library/digital-economy-and-society-index-desi-2022.

European Parliament (2020), Briefing. Artificial intelligence act and regulatory sandboxes, PE 733.544.

European Parliamentary Research Service (2020), Artificial intelligence: From ethics to policy, Scientific Foresight Unit (STOA), PE 641.507.

Floridi, L. (2020). Artificial Intelligence as a Public Service: Learning from Amsterdam and Helsinki. *Philos. Technol. 33*, 541–546.

Foth, Marcus. (2018). Participatory Urban Informatics: Towards Citizen-ability. *Smart and Sustainable Built Environment*, 7(1), 4-19.

Granlund, T. & Vedenpää, J. (2020). RegOps-diving into the dilemma of agile software development in regulated industry. [Online] https://www.solita.fi/en/blogs/regops-diving-into-the-dilemma-of-agile-software-development-in-regulated-industry/

Haataja M, van de Fliert L & Pasi R. (2020). Public AI Registers: Realising AI transparency and civic participation in government use of AI. Whitepaper Version 1.0, Saidot.ai.

High-Level Expert Group on Artificial Intelligence (2019). Ethics Guidelines for Trustworthy AI, European Commission, Brussels.

Hollands, R. G. (2008). Will the real smart city please stand up? Intelligent, progressive or entrepreneurial? *City 12, 3*: 303–320.

Jobin, A., Ienca, M. & Vayena, E. The global landscape of AI ethics guidelines. *Nat Mach Intell* **1**, 389–399 (2019). https://doi.org/10.1038/s42256-019-0088-2

Juuti, M. & Mattila, J. (2022). Machine Learning in Social Security — does it affect inclusion or equity?. [Online] https://medium.com/kelalab/machine-learning-in-social-security-a-case-study-3287da9a6c01

Kela (2019), Kelan tekoälyn eettiset periaatteet, Työpajakokonaisuus 2019.

Kitchin, R. (2016). The Ethics of Smart Cities and Urban Science. *Phil. Trans. R. Soc. A 374.*

Kitchin, R. (2019). Reframing, reimagining and remaking smart cities. *Creating smart cities.*

Koski, O. (2018). Work in the age of artificial intelligence: Four perspectives on the economy, employment, skills and ethics. *Ministry of Economic Affairs and Employment of Finland.*

Mattila, J. & Juuti M. (2022). Machine learning to identify customer needs, KelaLabs, Workshop on Inclusion, Equity and Accountability at the *Data, AI, and Public Sector Governance Conference*. [Online] https://www.datalit.fi/data-ai-and-public-sector-governance/

Manzoni, M., Medaglia, R., Tangi, L., Van Noordt, C., Vaccari, L. & Gattwinkel, D. (2022). AI Watch. Road to the adoption of Artificial Intelligence by the public sector. *Publications Office of the European Union, Luxembourg,* doi:10.2760/288757, JRC129100.

Medaglia, R., Misuraca, G. & Aquaro, V. (2021). Digital Government and the United Nations' Sustainable Development Goals: Towards an analytical framework. *DG.O2021: The 22nd Annual International Conference on Digital Government Research,* 473–478. https://doi.org/10.1145/3463677.346373

Mehr, H. (2017). Artificial Intelligence for Citizen Services and Government. *Harvard.* https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

Ministry of Economic Affairs and Employment of Finland (2021), Artificial Intelligence 4.0 First interim report: from launch to implementation stage, 2021:53.

Netherlands Court of Audit (2022). An audit of algorithms. Nine algorithms used by the Dutch government, *The Hague: Department of Communication.*

OECD (1999), "European Principles for Public Administration", *SIGMA Papers,* No. 27, OECD Publishing, Paris, https://doi.org/10.1787/5kml60zwdr7h-en.

OECD (2021). Artificial Intelligence: Regulation Can Support Innovation (OECD Podcasts).

Pechtor, V. & Basl, J. (2022). Analysis of suitable frameworks for artificial intelligence adoption in the public sector, *IDIMT-2022 Digitalization of society, business and management in a pandemic: 30th Interdisciplinary Information Management Talk*, doi: 10.35011/IDIMT-2022-67.

Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual liability rules to artificial intelligence (AI Liability Directive), COM(2022) 496 final, Brussels, 28.9.2022.

Proposal for a Regulation Of The European Parliament And Of The Council Laying Down Harmonised Rules On Artifcial Intelligence (Artifcial Intelligence Act) And Amending Certain Union Legislative Acts, COM/2021/206 fnal, Brussels, 21.4.2021.

Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act), COMM(2022) 767 final, Brussels, 25.11.2020.

Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act), COM(2022) 68 final, Brussels, 23.2.2022.

ProPublica (2016). Machine Bias. [online] https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Pūraitė, A., Zuzevičiūtė, V., Bereikienė, D., Skrypko, T. & Shmorgun, L. (2020). Algorithmic Governance in Public Sector: Is Digitization a Key to Effective Management, *Independent Journal of Management and Production (IJM&P),* 11(9), pp. 2149-2170.

Ranawana, R. & Karunananda, A. S. (2021). An Agile Software Development Life Cycle Model for Machine Learning Application Development, *5th International Conference on Artificial Intelligence*, doi:10.1109/SLAAI-ICAI54477.2021.9664736.

Regulation (EU) 2016/670 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC.

Ries, J., Santo, P. E. & Melão, N. (2019). Impacts of Artificial Intelligence on Public Administration: A Systematic Literature Review, pp. 1-7.

Rolfe, C. (2019) Citizens' Juries: Using public opinion on Artificial Intelligence to inform policy. Royal Society for the encouragement of Arts, Manufactures and Commerce (RSA) (2019), Democratising decisions about technology: A toolkit. [Online] https://bit.ly/3F4t7e6.

Ruckenstein, M., Lomborg, S. & Hansen, S. (2020) Re-humanising automated decision-making. Workshop report from the *ADM: Nordic Perspectives research network.*

Ruf, P., Madan, M., Reich, C. & Ould-Abdeslam, D. (2021). Demystifying MLOps and Presenting a Recipe for the Selection of Open-Source Tools. *Applied Sciences*, 11(8861).

Saidot (2022). Introducing Saidot Policy Template for AI Deployers. [Online] https://www.saidot.ai/post/introducing-saidot-policy-template-for-ai-deployers.

Sawhney N. & Anh-Ton T. (2020) Ecologies of contestation in participatory design. *In: Proceedings of the 16th Participatory Design Conference (PDC 2020)*, June 15–19, 2020, Manizales, Columbia. ACM.

Sawhney, N. Contestations in urban mobility: rights, risks, and responsibilities for Urban AI. *AI & Soc* (2022). https://doi.org/10.1007/s00146-022-01502-2

Schaefer, C., Lemmer, K., Kret, K. S. (2021), Truth or Dare? – How can we Influence the Adoption of Artificial Intelligence in Municipalities?, *Proceedings of the 54th Hawaii International Conference on System Sciences*, p. 2347-2356.

Stirbu, V., Granlund, T. & Mikkonen, T. (2022), Continuous design for machine learning in certified medical systems, *Software Quality Journal,* http://dx.doi.org/10.2139/ssrn.3064761

Truby, J., Brown, R., Ibrahim, I., & Parellada, O. (2022). A Sandbox Approach to Regulating High-Risk Artificial Intelligence Applications. *European Journal of Risk Regulation, 13*(2), 270-294. doi:10.1017/err.2021.52

von der Leyen, U. A Union that strives for more. My agenda for Europe. Political Guidelines for the Next European Commission 2019-2024.

Yfantis, V. & Ntalianis, K. (2019). Exploring the Adoption of the Artificial Intelligence in the Public Sector. *International Journal of Machine Learning and Networked Collaborative Engineering*, 3(4), pp. 210-218.