The Review of Socionetwork Strategies (2023) 17:297–318
https://doi.org/10.1007/s12626-023-00146-y

ARTICLE

Role of Regulatory Sandboxes and MLOps for AI-Enabled Public Sector Services

Ana Paula Gonzalez Torres[1] · **Nitin Sawhney**[1]

Received: 6 March 2023 / Accepted: 1 August 2023 / Published online: 18 August 2023

Abstract

This paper discusses how innovations in public sector AI-based services must comply with the Artificial Intelligence Act (AI Act) regulatory frameworks while enabling experimentation and participation of diverse stakeholders throughout the *Artificial Intelligence (AI) lifecycle*. The paper examines the implications of the emerging regulation, *AI regulatory sandboxes* and *Machine Learning Operations* (MLOps) as tools that facilitate compliance while enabling co-learning and active participation of multiple stakeholders. We propose a framework that fosters experimentation with automation pipelines and continuous monitoring for the deployment of future public sector AI-based services in a regulatory-compliant and technically innovative manner. AI regulatory sandboxes can be beneficial as a space for contained experimentation that goes beyond regulatory considerations to specific experimentation with the implementation of ML frameworks. While the paper presents a framework based on emerging regulations, tools and practices pertaining to the responsible use of AI, this must be validated through pilot experimentation with public and private stakeholders and regulators in different areas of high-risk AI-based services.

Keywords Artificial intelligence · Public sector · AI Act · Regulatory sandboxes · Machine learning operations (MLOps) · Multi-stakeholders

1 Introduction

Public sector organisations are increasingly embracing algorithmic decision-making and data-centric infrastructures in their efforts to incorporate artificial intelligence (AI)-based systems to improve digital services in areas, such as education,

✉ Ana Paula Gonzalez Torres
ana.gonzaleztorres@aalto.fi

Nitin Sawhney
nitin.sawhney@aalto.fi

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

healthcare, and urban mobility [1–3]. Consequently, this leads to challenges such as designing durable systems compliant with emerging regulatory measures and involving multiple stakeholders for enhanced trustworthiness [4, 5].

This trend towards leveraging AI-based services is motivated by the public sector's essential goal of better serving its recipients (citizens and non-citizens alike) in a fair, timely, and cost-effective manner [6]. The use of big data, algorithms, and machine learning to develop AI-enabled systems is often seen as a way to improve such digital public services [7, 8]. Innovations in *AI-enabled public services* often 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) [6, 9]. For instance, public authorities' use of 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, as well as gather feedback on the design of public sector services [10]. In local environments, urban mobility data pooled from diverse sources (using historical and real-time data) can contribute to better urban planning and modeling, with AI algorithms learning and predicting citizens' patterns of movement for participatory design of suitable transportation alternatives and pathways [12].

As AI-enabled public services become more prevalent, they impact people's lived experiences [11]. Examples include the unforeseen use of biased algorithms in criminal risk assessment in the U.S. [12], algorithmic systems that downgraded less privileged student's examination scores in the U.K. [8], 'discriminatory algorithms' used by the Dutch Tax Authority for the provision of childcare benefits [13], and unlawful debt schemes that have affected already vulnerable people in Australia [14]. In light of such potential adverse impacts, there has been a call for regulatory measures [5] that establish the rights, risks, and responsibilities of people involved in the AI value chain, including recipients of such AI-enabled services. It has resulted in the European Commission's proposed "Artificial Intelligence Act" (EC AI Act) [15]. At the time of writing, there are different versions of the AI Act which are meant to be the basis for the legislative trialogue between the European Commission (EC AIA), Council of the European Union (CE AIA), and European Parliament (EP AIA). However, it is the diverse and contested discourses around the AI Act among legal experts, regulators, as well as private and public actors which offer a timely window of opportunity to critically examine the challenges and concerns emerging from it [16, 17].

The AI Act introduces new considerations and challenges as its risk-based approach sheds light on public services as potentially "high-risk" and thus falls within its regulatory scope. The first challenge for the public sector is fostering innovation in digital services while having to comply with the emerging regulatory framework of the AI Act; we believe that this necessitates the use of AI regulatory sandboxes for regulatory experimentation. The second challenge is the need to ensure AI systems follow a "lifecycle" approach with ongoing monitoring and auditing [15]; we believe that this necessitates the use of technological processes such as Machine Learning Operations (MLOps). A final challenge comes from the multiple stakeholders involved in public sector services and emerging regulatory requirements to consider diverse stakeholder participation as relevant sources for feedback through the lifecycle of AI systems [18].

In the following sections, we discuss how potential AI-based public sector services would be impacted by the proposed AI Act and the need to foster innovative experimentation in technical processes and regulation through multi-stakeholder participation throughout the AI lifecycle; in particular, we examine AI-enabled services in the public sector.[1] We believe that it is crucial to facilitate experimentation and co-learning in deploying future "high-risk" AI-enabled public sector services. In this paper, we examine the implications of the AI Act for AI-enabled public sector services, review prior literature on regulatory sandboxes and MLOps, and examine how they can be leveraged for responsible development, regulation, compliance, and governance of such systems. Such efforts must align with the goal of "advancing regulation through proactive regulatory learning and enabling regulators to gain better regulatory knowledge" at an early stage of development amid high risk, uncertainty, and disruptive challenges, as per the Council of the European Parliament document 13026/20. Any framework proposed for AI regulatory sandboxes must be validated through pilot experimentation with multiple stakeholders from the public and private sectors as well as regulators in different domains of high-risk AI-enabled services.

2 Implications of the AI Act in Public Sector Services

In the EC AI Act of April 2021, the explanatory memorandum proposes a "risk-based regulatory approach" which, according to article 1(a), aims to lay down "harmonised rules for the placing on the market, the putting into service and the use of artificial intelligence systems ('AI systems') in the Union". It was followed in November 2022 by a general approach version by the Council of the European Union and a compromise draft by the European Parliament in May 2023. We recognise that at the time of writing, there are still vigorous debates and legislative deliberations emerging that will continue to shape the proposed AI Act and its compliance framework in the European Union (EU). Nonetheless, there is a timely opportunity to examine the emerging legislative measures and public deliberations around the AI Act as they contribute to the formation of emerging legislation.

The AI Act has important implications for future AI-enabled public sector services. Remarkably, all three versions of the AI Act, in Annex III point 5, categorise as "high-risk" the AI systems employed in the area of "access to an enjoyment of essential private services and public services and benefits" [19]. Thus, there is a latent potential for AI-enabled systems used in the public sector to be characterised as "high-risk" and come under the scope of the AI Act. Should the proposal become law, governments and municipalities must assess and adapt their AI-enabled services to comply with this new regulation while promoting innovation and multi-stakeholder participation [14, 15]. In practice, once categorised as "high-risk", AI

[1] The idea for this paper stems from a previous conference presentation: Sawhney, N. & Gonzalez Torres, A. P., 2022, Devising Regulatory Sandboxes and Responsible Practices for Designing AI-based Services in the Finnish Public Sector, International Workshop on AI Compliance Mechanism (WAICOM 2022), 35th International Conference on Legal Knowledge and Information Systems (JURIX 2022), Saarland University, Saarbrücken, Germany.

systems would have to comply with therein established regulatory requirements such as:

- Implementation of risk management system through the entire lifecycle (article 9);
- Quality dataset and data governance (article 10);
- Technical documentation demonstrating conformity to the requirements (article 11);
- Record keeping enabling for automatic recording of events in a traceable and monitoring fashion (article 12);
- Transparency and provision of information to users are similar to instructions for use (article 13);
- Human oversight in a manner that enables reduction of risks and monitors operations (article 14);
- Accuracy, robustness, and cybersecurity for the entire lifecycle (article 15).

As a complex legislation, the AI Act must grapple with technical details of a continually evolving scientific domain (artificial intelligence) and technological innovations (like ChatGPT) in a future-proof manner that nonetheless considers established legal principles. Hence, public sector organisations must engage with responsible AI practices throughout the AI development and deployment lifecycle. For instance, in the conceptualisation stage, to understand whether there is a real need for an AI-based approach in offering a particular public service [20]; in the training of datasets, ensuring that the data are unbiased [21, 22]; in the development stage whether there is appropriate verification, testing, and validation of outcomes [23]; in the deployment stage what harmful or unforeseen impacts may emerge [24]; during maintenance, whether the use of an AI-based system creates discriminatory feedback loops [11]; and finally during retirement, considering implications of recalling a system that people have come to rely on.

In legislative terms, the architecture of the AI Act is based on the New Legislative Framework, a typical product safety approach, based on the notion that the manufacturer (considered as a provider in the AI Act) has detailed knowledge of the design and production process and thus is best placed to carry out conformity assessment procedures [25]. Nonetheless, for public sector services, the complexity of the public sector also comes from the involvement of different stakeholders throughout the AI lifecycle. As such, while public sector organisations could be involved in designing AI-enabled services, a third party may supply data for AI training [26]. Meanwhile, development may be contracted to a private sector organisation. Thus, to embed regulatory considerations in public sector AI practices, each stakeholder involved in the process would bear certain roles and responsibilities at nearly every stage of the AI lifecycle in an iterative manner.

From this legal complexity derives the consensus in different documents presented during the feedback period to the proposed AI Act[2] that the AI Act could

[2] https://digital-strategy.ec.europa.eu/en/library/impact-assessment-regulation-artificial-intelligence.

hinder innovation and future technological development in Europe. The implication for public sector services is that a complex regulatory regime such as the AI Act would make some services inflexible for digitalisation and data utilisation.[3] To address such concerns, the AI Act proposes the development of AI regulatory sandboxes (Title V, EU Regulation Proposal). This presents a timely opportunity to leverage them as "measures in support of innovation" for experimentation with "a view to ensuring compliance with the requirements of this [AI Act] Regulation" (article 53). In the public sector, institutions' responsibility towards recipients calls for strict legal compliance that supports societal needs [27]. Thus, developing an understanding of the inter-relations between the different regulatory measures would be better served by an experimental environment that facilitates proactive oversight of the development of AI-enabled public sector services and potential adaptation of relevant laws by regulatory authorities rather than the imposition of fines or other punitive measures to ensure compliance.

3 Adoption of AI-enabled Systems by the Public Sector

Adopting AI-enabled systems in the public sector requires an AI governance framework that considers a broader ecosystem of stakeholders, from developers, providers, and regulators to end-users [11]. Such an AI governance model should establish the roles, legal requirements, and technical measures related to AI systems and their practical enforcement and application. For instance, an effective AI governance model should incorporate an iterative and continuous development process covering the whole AI lifecycle from design, development, deployment, and operational usage to inevitable retirement [28, 29]. This necessitates leveraging frameworks like MLOps, a set of technologies and practices for continuous deployment, maintenance, and monitoring of machine learning models [30].

Such AI governance should guide the development of AI-enabled services in the public sector [23], underpinning the mandate of the public sector for equitable, fair, and inclusive services [31]. The public sector must adhere to the *rule of law* as a legal principle that dictates what and how it pursues the goals of public good for its diverse stakeholders, for example, ensuring scrutiny and accountability, applying equity, transparency, and consistency in decisions and redress whenever needed [9]. Amnesty International (2021) states that "governments around the world are rushing to automate the delivery of public services, but it is the most marginalised in society that are paying the highest price" [32].

AI-enabled systems in the public sector are often developed across a complex multi-stakeholder ecosystem with some aspects devised in-house, others procured as

[3] Not to mention that the public sector must also contend with the emerging 'Data Act', 'Data Governance Act', 'AI Liability Directive', and local legislation on how certain services must be organised and provided in practice. In this paper, we will focus on the AI Act but acknowledge the complexities of the legal system that underpin local AI innovations which are outside the scope of this paper but need to be taken in consideration in public sector innovation.

software or services from the private sector, or the result of public–private partnerships for co-development [33]; each of these impose different sets of requirements and obligations under the regulatory measures that pertain to each actor. Furthermore, there is often poor coordination between different public administration bodies as they often work in silos [34]. Thus, a lack of communication or integration between units experimenting with AI-based systems 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 [35].

These challenging conditions require novel approaches to support public sector AI innovation. We examine how these regulatory AI sandboxes can be devised with MLOps-based processes to facilitate innovative AI services in the public sector. There is a need to provide an experimental space for regulatory and technological innovation, particularly in high-risk domains. MLOps frameworks can offer agile and dynamic tools for technical and responsible adoption across the AI lifecycle in the public sector [36]. Such frameworks can also assist managers and developers of public services in piloting and technically validating them before they are launched into production. Regulatory sandboxes in other sectors like financial services have been used to explore the implications of using 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 mediating potential risks on a larger scale [9].

4 Experimentation using AI Regulatory Sandboxes and MLOps in the Public Sector

It has been recognised that the inclusion of experimental regulatory instruments in the proposed AI Act can be partially explained by the need to accommodate future developments in AI and address their inherent complexity [37]. New technologies require regulators to make several complex decisions regarding regulation and measures to support their uptake. Especially in the public sector, the AI Act should also uphold and support the public sector organisations' participation in AI regulatory sandboxes. Thus, in the proposed framework, we consider AI regulatory sandboxes as spaces for experimentation to understand whether integrating MLOps processes can aid with compliance efforts and active engagement of multiple stakeholders throughout the AI lifecycle in public sector services. We acknowledge the existence of various mechanisms designed to support and ensure compliance, such as human rights due diligence, impact assessment, certification and standards, auditing and monitoring, and regulatory sandboxes [38]. Even though these mechanisms promote best practices, such as the reflective and anticipatory assessment of an AI-based system, the use of compliance should evolve from the earliest stages of project design to ongoing mechanisms for monitoring the system following its deployment to account for any changes in its function [mlops.org]. Examples of types of compliance mechanisms depend on the context (e.g., different regulatory cultures) and the diversity of components of an AI system subject to compliance (e.g., training data features). To help determine the mechanisms best suited to each context, inclusive and participatory processes should be carried out with relevant stakeholders. Dynamic (not

static) assessment at the beginning and throughout the AI project lifecycle can be used to account for ongoing decision-making. Thus, our envisaged framework combines interdisciplinary research in mechanisms that could allow for a technology-adaptive approach and support efforts at futureproofing.

4.1 Experimental Legislation: AI Regulatory Sandboxes

Regulatory sandboxes are considered policy instruments usually used within the concept of advisory and adaptive regulation [39]. They initially emerged in the FinTech sector, introduced by UK's former chief scientific adviser Sir Mark Walport, as "testing ground" tools for regulatory innovation, providing a balance between supporting innovation and creating regulatory measures [40]. Sandbox testing can help participants gather valuable regulatory input in the design, evaluate strategy, and devise potentially "safer" product features, as risk assessment and technical requirements, such as accuracy and performance, can sometimes only be established in a real-life setting [41]. The main characteristics are the critical role of regulating authorities, the involvement of governmental actors, either local or national, and the co-learning process [42]. It requires an open attitude, proper set-up of the experiment, continuous monitoring, and stringent evaluation to maximise value knowledge gain [43].

As part of the legal system, experimental frameworks such as regulatory sandboxes must abide by well-established principles such as legal certainty and equal treatment. Thus, they are to be designed clearly and objectively, not contrary to the principle of legal certainty, and to prevent situations where citizens do not know the content of enforced laws. Nonetheless, outdated laws that do not account for societal changes violate the principle of legal certainty [37]. In terms of equal treatment and market actors, because of the risk of regulatory distortion that could affect competition in the market, legislative experimentation only to a limited number of its potential subjects is compatible with such principles as long as experimental laws have a transitory character and the trial takes place according to objective criteria [44].

The EC proposed AI Act article 53(1) establishes "AI regulatory sandboxes" as 'controlled environments" within a pre-market deployment phase in the AI lifecycle under the "direct supervision and guidance of competent authorities" for a limited period of time. They would thus allow technical learning from development, testing and validation of AI systems in a "real-world environment", as well as legal experimentation with regulatory regimes [9, 26]. According to the Internal Market and Consumer Protection-Civil Liberties, Justice and Home Affairs proposed amendment to Article 53, "5 a. Regulatory sandboxes shall allow and facilitate the testing of possible adaptations of the regulatory framework governing artificial intelligence to enhance innovation or reduce compliance costs, without prejudice to the provisions of this Regulation or the health, safety, fundamental rights of natural persons or to the values of the Union as enshrined in Article 2 TEU. The results and lessons learned from such tests shall be submitted to the AI Office and the Commission." [45].

Here cooperation between multiple stakeholders within AI regulatory sandboxes influences the experiment's outcome as it depends on who participates [46]. While experimenting with regulation directly affects the regulator and regulated, it is critical to engage with a broader group of actors, including those who would be indirectly impacted [43]. This is corroborated by the draft of the European Parliament, which is an added article 53a, establishes that "regulatory sandboxes facilitate the involvement of other relevant actors within the AI ecosystem, such as notified bodies and standardisation organisations (SMEs, start-ups, enterprises, innovators, testing and experimentation facilities, research and experimentation labs and digital innovation hubs, centers of excellence, individual researchers), in order to allow and facilitate cooperation with the public and private sector".

The main benefit of an AI regulatory sandbox within our proposed framework is the possibility of determining the effectiveness and feasibility of AI requirements. We may learn valuable lessons about their practical applications by "piloting" associated tools and templates in the sandbox [47]. One of the critiques in the literature is that regulatory sandboxes have not offered truly novel regulatory responses to traditional regulation. Instead, they repurpose old technocratic tools to fill specific regulatory gaps [48]. Hence, the importance of our work as it combines regulatory sandboxes with technical capabilities as a tool to comply with the regulation in a continuous monitoring fashion, which could potentially enable different stakeholders, from regulators to end-users, to participate in lawful AI systems through its lifecycle actively.

From the regulator's perspective, regulatory sandboxes can afford them a better understanding of the product or service. It could lower barriers to innovation as legal uncertainty can implicitly lead to projects being abandoned at an early stage or never undergoing testing and validation [46, 49, 50]. For the public sector, regulatory sandboxes can allow more innovative products to reach citizens, as they facilitate co-learning about risks during the testing stage and can help reduce the time-to-market with lower costs for the organisation in verifying and demonstrating the success of technological innovations [38, 45]. Consequently, more AI-enabled public services could be tried and later introduced to the market or put into service upon validation [46] and in a long-lasting manner if implemented utilising MLOps principles, as we will explore in the next section. For the different stakeholders, participation in a regulatory sandbox means adding another layer of regulatory supervision which should be carefully examined as it has been recognised that experimental legislation while trying to reduce the individual burdens for individuals has also increased the overall number of regulatory burdens as experimental regulations also establish new compliance rules [37]. Furthermore, it requires investing economic and human resources as the experiment requires setting up, running, and being prepared to actively contribute to the sandbox in different phases [51].

A criticism of the proposed AI Act is that participants in AI regulatory sandboxes would remain liable for any harm inflicted on third parties due to experimentation in the sandbox environment. Nonetheless, within our proposed framework, we envision that MLOps capabilities for monitoring, versioning, and documenting events can facilitate the tracing and allocation of potential liability as errors that pose a risk could be flagged and addressed during experimentation [29]. This can implicitly

reduce the risk of harm (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 in terms of duty of care and defences to negligence.

There are two limitations to our proposed framework. First, as it has been raised because of the legislative state of the AI Act, it is unclear what the limitations of the future AI sandboxes will be, what type of regulatory relief they are allowed to provide, and how they will be funded [37]. Nonetheless, we acknowledge the current efforts in Spain for the first AI regulatory sandbox and thus expect more clarity shortly as "results will be published during the Spanish Presidency of the Council of the EU in the second half of 2023."[4] Second, even if, according to the EC AI Act, regulatory sandboxes could extend "where relevant, [to] other Union and Member States legislation supervised within the sandbox", we will only focus on compliance measures with the AI Act this is justified by the acknowledgement that experimentation within AI regulatory sandboxes "shall not affect the supervisory and corrective powers of competent authorities" ex article 53(3). Thus, personal data protection will fall outside the scope of this paper as there are other examples from which lessons can be learned, such as the Norwegian sandbox, which focuses on data protection.[5]

While sandboxes can be an important tool for developing evidence-based policies, they are not to be considered as the all-encompassing solution to AI-based systems compliance but a tool to better inform compliance measures and regulatory intervention.[6] Nonetheless, when specifically devised for experimentation with AI-based systems and services, in conjunction with automation and monitoring via software frameworks like MLOps, AI sandboxes lead to more significant opportunities and benefits for deploying regulatory AI sandboxes.

4.2 A Software Development Framework: Machine Learning Operations (MLOps)

One key feature of an ML system is the ability to learn, adapt, and optimise operations in real time. The ability to improve by learning from data while performing critical operations may be the most valuable asset of ML technology. Systems that incorporate AI as machine learning (ML) technology provide the possibility to improve services such as (i) aggregating and analysing information from multiple sources to the extent that would not be possible without ML, (ii) personalisation and trend modeling to serve patterns and interests of different user groups, for example by clustering customer data and fitting an ML model separately; (iii) automation and

[4] https://digital-strategy.ec.europa.eu/en/news/first-regulatory-sandbox-artificial-intelligence-presented.

[5] https://www.datatilsynet.no/en/regulations-and-tools/sandbox-for-artificial-intelligence/.

[6] For example, the Banca Negra Malaysia (BNM) experience with UK-based remittance company, WorldRemit, resulted in the bank amending its electronic know your customer (eKYC) regulations to permit remittance providers to verify customer identities via "selfie" and other remote identifiers. *See* UNSGSA FinTech Working Group and CCAF. (2019). Early Lessons on Regulatory Innovations to Enable Inclusive FinTech: Innovation Offices, Regulatory Sandboxes, and RegTech. Office of the UNSGSA and CCAF: New York, NY and Cambridge, UK.

self-service, whereby chatbots and online recommendation systems serve customers on-demand [23].

Regarding regulatory compliance, MLOps approach supports automation, continuous monitoring, documentation, traceability, and auditing of the resulting AI systems, which can be done by multiple tools as follows [29]. Automated pipelines could provide a feedback loop to each stakeholder from all the process stages. As the software progresses through the pipeline, different stages can be triggered. For example, metadata documentation of the historical sequence of key metrics (e.g., data quality of test data, reliability) can be viewed manually by queries or by automatic generation of documents. Traceability by capabilities such as "running issues" can be used during implementation to track tasks or bugs and to plan future work by collecting ideas and feedback [35, 52]. Continuous monitoring ensures that the model behaves as expected and that anomalies are detected and addressed correctly [53]. Continuous evaluation/audit-based quality assurance based on automatically determined key metrics can make information collection independent of manual collection [53].

As a warning, although MLOps lean on continuous improvement, it is often a complex process involving changes in the application code, the model used to provide an outcome (e.g., for prediction), and the data used to develop the model [54]. It could indirectly affect compliance with regulatory requirements. As an example, EC AI Act requirement on 'data and data governance' established that training, validation, and testing data 'shall have the appropriate statistical properties, including, where applicable, as regards the persons or groups of persons on which the high-risk AI system is intended to be used' (article 10) or that technical documentation "of a high-risk AI system shall be drawn up before that system is placed on the market or put into service and shall be kept up-to-date" (article 11). Furthermore, an ML system's autonomous operation raises concerns about safety in highly regulated sectors (e.g., clinical performance in the medical device domain). Authorities have traditionally promoted the approach of "locked" algorithms, where the system is designed so that it is being trained during the development phase, and the ability to improve is disabled in real-world use [54].

In terms of the public sector, data products are highly sensitive and require constant monitoring, repairing, and updating [23]; thus, they could benefit from employing MLOps framework for durable AI-based innovation in public services. AI regulatory sandboxes and MLOps support more proactive and scalable exploration of innovative AI-based services that would otherwise not be attempted because of their categorisation as "high-risk"; thus, long-term AI innovation with higher impact requires dynamic means of integrating compliant practices in the AI development lifecycle. In the following Sect. 5.3 we outline in more detail how various MLOps methods could be utilised in AI regulatory sandboxes.

5 Framework: Multi-stakeholder AI Regulatory Sandboxes & MLOps

In this section, we will describe the proposed framework for multi-stakeholder AI regulatory sandboxes and the instrumentalization of MLOps for the development of "high-risk" AI systems in the public sector. Even though the proposed AI Act specifies privileged access to SMEs, based on the understanding that the public sector has potential high-risk AI-enabled services, it would be beneficial to provide regulatory sandboxes experimentation to public sector services which intend to further public good rather than a competitive advantage in the marketplace. Furthermore, sandboxes can be viewed as a tool for managing regulatory fears by developers or providers of such AI-based systems or services (e.g., by taking a data-driven approach for fair and proactive guidance when approving pilots or issuing regulatory guidance). It is meant to support limited access to real-world user and regulatory environments, with the participation of stakeholders in a virtual pilot context. A virtual sandbox could provide an environment that enables organisations to validate their systems and services in a contained virtual space without placing them in the market or wider deployment with all end-users [54]. Participation in AI regulatory sandboxes would be most beneficial to public sector organisations in the early stage of implementing AI-based systems or services considered high-risk according to the AI Act.

Based on an analysis of the three versions of the AI Act by the European Commission (EC AIA) [15], the Council of the European Union [55] and the European Parliament [18], we identified several characteristics which can be clustered within the three main phases of regulatory sandboxes: (1) preparation and planning; (2) legal aspects; (3) design and implementation [27].

5.1 Preparation and planning

The first phase of preparation and planning requires designing and utilising networks to establish the involvement of stakeholders, planning time, and resources while being aware that the experiment's design can wrongfully capture an innovation's potentially problematic effects [37, 48]. Within the AI lifecycle, it is the conceptualisation stage. There is a need for a participatory design that permits a collaborative approach. The EC AIA has depicted the general structure of the sandbox as an environment "established and guided by member states competent authorities or the European Data Protection Supervisor". At the same time, the EP draft includes the "possibility for regional/local or jointly with other member states".

The proposed multi-stakeholder AI regulatory sandbox would need to initially establish the stakeholders that could participate in the sandbox based on their roles as stakeholders. They can be included as core stakeholders, active participants, occasional participants, or part of the surrounding environment. In Fig. 1, the AI regulatory sandbox depicts the virtual space in which the AI system is being tried as an MLOps implementation with the inclusion of multiple stakeholders. The involvement of the different stakeholders is aided by the ML pipelines, which consist of tools that support the process, from code to delivery. These tools ensure that each stakeholder gets timely access to what they need [54]. Finally, an interface and

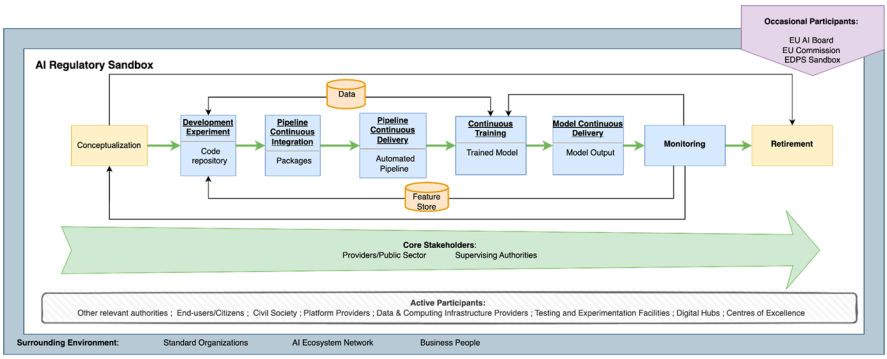

Fig. 1 Multi-stakeholder AI Regulatory Sandbox implementing MLOps

credentials (e.g., based on their role within the sandbox) should give access to the model pipeline to participate.

Core stakeholders *decide* in a participatory manner on the set-up, design of the regulatory sandbox, and implementation of subprojects. It includes stakeholders with high decision-making powers, such as member state competent authorities or European Data Protection Supervisor (EDPS), which would have a supervising role and provider/public sector wishing to experiment with AI-based public services. They are included in every stage of the pipeline to pursue the implementation of the regulatory sandbox in their interests, with open and regular dialogue (e.g., monthly workshops) and a critical view of the project. In a multi-stakeholder participatory environment, core stakeholders should be provided with credentials to access the MLOps pipeline under different levels of access and intervention to gather relevant information and trigger specific actions. On the one hand, supervising authorities allow and guide prospective providers to fulfil, in a controlled environment, the conformity assessment obligations of this regulation or voluntary application of the codes of conduct. Based on the EP AIA draft, "guide and supervision within the sandbox to identify risks, in particular to fundamental rights, democracy, and the rule of law, health and safety and the environment" and "on how to fulfil the requirements of the AIA so that the AI system may exit the sandbox being in the presumption of conformity with the specific requirements of the regulation that were assessed within the regulation, in so far as it complies with the requirements when exiting the sandbox". On the other hand, provider/public sector organisation efforts in compliance would thus include an active and continuous process of documentation that covers the whole life cycle allowing the mapping of relevant information and comparably evaluating the state of implementation.

Active participants *supply* services to fulfil tasks and key components of innovation or to meet regulatory conditions (e.g., under AI Act article 15, they participate in providing cybersecurity solutions). They have access to and influence on core stakeholders at different phases of the AI regulatory sandbox and AI lifecycle. This role's stakeholders include other relevant authorities, end-users/citizens, civil society, partnering platform providers, data & computational infrastructure providers,

Testing and Experimentation Facilities, Digital Hubs, and Centres of Excellence. End-users/citizens could be involved in the first phase of the sandbox participating in the conceptualisation of what the AI-based service should be, and also at the sandbox phase of implementation as they could be involved by being end-users of the AI-based service that is being monitored based on their interactions. For instance, AI-enabled services which provide intermediate access to essential public services related to employment or health would need to involve appropriate stakeholders and regulators related to those domains. Regarding experimentation in a safe space, the sandboxes can offer limited participation of citizens who consent. The role of active participants like end-users/citizens is vital as they can perform an evaluation based on the key figures provided by the manufacturer and their lived experiences. Citizens should also be provided feedback channels and included in validating new digital services.

Occasional participants *support* by providing access to relevant networks, positive support with expertise, interpretation of legislation, and funding. They are door-openers to key authorities, can influence the public image, and foster/impede successful implementation. This stakeholder category can include the EU Artificial Intelligence Board, EU Commission and EDPS. They are to be kept updated about the progress made and—if necessary—explicitly addressed when there is a need for their input in the governance of AI. Based on the EC AIA draft, there needs to be "coordination and cooperation with the European Artificial Intelligence Board" and "where appropriate other national authorities & other actors in the AI ecosystem", as the EP AIA draft adds. In terms of the EC, the EP AIA draft adds that it should "establish a dedicated interface containing all relevant information related to sandboxes and allow stakeholders to raise enquiries with competent authorities and seek non-binding guidance on the conformity of innovative products, services, business models embedding AI technologies". The EC could have a "complementary role enabling Member States to build on their expertise and assisting and providing technical understanding and resources to those Member States that seek guidance on the set-up and running of these regulatory sandboxes" as stated by the EP AIA draft.

Surrounding environment *observes*, may participate, possibly passively, and is indirectly or in any way impacted by the results of the AI regulatory sandbox. This stakeholder group can include standardisation organisations, AI ecosystem networks, residents, or businesses. They are essential for public acceptance of the experimentation at an early stage. Thus, individual groups or the general public should be informed transparently and openly and allowed to transition to the role of active participants in the decisions and processes.

In this first phase, the duration of the experimentation and the place must be chosen. It could be a virtual space as it allows flexibility in gathering multiple stakeholders' inputs regarding compliance efforts. In terms of duration, it should be taken into consideration that specific experiments may deliver immediate results (e.g., direct complaints resulting from direct discrimination by AI applications), and others may require more time to show outcomes (e.g., indirect long-term discrimination by sophisticated AI applications) [37]. Experimentation is supposed to occur for a limited time before market placement, or it is put into service; nonetheless, as appropriate based on the complexity and scale of the project, national competent

authorities could offer extension [55]. As a complement to the automation of the pipelines, workshops can be established every month, depending on the experimentation's duration, to ensure there is co-learning and different stakeholders can share their experiences simultaneously. It can also better understand end-users'/citizens' experiences [56].

The allocation of resources should be considered by all willing participant stakeholders, as costs could prove to be an obstacle if they are not foreseen. For instance, in an autonomous delivery robot case, the cost of participating in a 7-month test amounted to €100,000 for the applicant. Similarly, to support innovative businesses in Germany via "regulatory innovation zones", Economic Affairs Ministry and Federal Network Agency faced one-off compliance costs of €60,000 and ongoing compliance costs of €69,000 for the notification and application procedures. "Most of the costs arise in the procedures introduced to offset disadvantages; these procedures are intended to encourage the participation of stakeholders who would otherwise not be interested" [42].

In the initial phase, the conditions for the operation of the AI regulatory sandbox must be established. According to the EC AIA proposal, "eligibility criteria, procedure for the application, selection, participation and existing from the sandbox, rights and obligations of the participants in implementing acts". Meanwhile, the CE AIA approach mentions that the "access to AI regulatory sandbox is to any prospective/provider that fulfils selection criteria & has been selected by competent national authorities following the selection procedure". In Norway, the regulatory sandbox has a committee with lawyers and experts that examine the feasibility and merit of applicants, which applies based on four criteria[7]: that their project (1) makes use of artificial intelligence or otherwise touch artificial intelligence; (2) benefit the individual or society; (3) could benefit significantly from participation in the sandbox; (4) be subject to the Norwegian Data Protection Authority as the competent supervisory authority. This could be implemented with proper modifications as the application criteria for AI regulatory sandboxes. The AI Act emphasises access to SMEs; such privilege access should also be extended to public sector innovative experiments, which could be presented in partnership with SMEs and other private business organisations. The CE AIA approach establishes that the "modalities and conditions shall to the best extent possible support flexibility for competent national authorities to establish and operate their AI regulatory sandboxes", and for the EP AIA draft, they should "ensure broad and equal access" also to be free of charge for SMEs and start-ups.

Upon admission to the AI regulatory sandbox, core stakeholders are to agree on a specific plan which defines goals and metrics to be evaluated during the experimentation [15, 55]. Because the purpose of an AI regulatory sandbox is to ensure compliance with the requirements of the AI Act for high-risk AI systems [15], this space shall aim to contribute to (a) foster innovation and competitiveness and facilitate the development of an AI ecosystem; (b) facilitate and accelerate access to Union

[7] https://www.datatilsynet.no/en/regulations-and-tools/sandbox-for-artificial-intelligence/framework-for-the-regulatory-sandbox/general-participation-criteria/.

Market for AI system, especially for SMEs; (c) improve legal certainty & development of best practices; (d) contribute to evidence learning regulatory learning [55]. For instance, the provider/public sector would indicate the requirements of the AI Act in which compliance experimentation would be most beneficial, because sandboxes shall facilitate the development of tools and infrastructure for testing, benchmarking, assessing and explaining dimensions of AI systems relevant to sandboxes, such as accuracy, robustness, and cybersecurity, as well as minimisation of risks to fundamental rights, environment and the society at large. They would also need to commit the necessary capacity, devoting relevant personnel and resources to actively participate in sandbox experimentation (e.g., data protection officers/procurers, data scientists, designers, and developers). Also, a nomination as an "ombudsman" for the AI regulatory sandbox experimentation to ensure critical reporting and oversight. Also, submit a contingency plan for citizen impact in case of failure/harm and establish the citizens who would be active participants.

5.2 Legal Aspects

The second phase of the AI regulatory sandbox relates to legal aspects. It requires exploring the legal obstacles and potential for regulatory manoeuvrings, such as granting exceptions. Regarding legal obstacles, fundamental laws and regulations besides the AI Act must be specified based on an impact assessment depending on the use case of the AI-based service. It is determining applicable laws and legal constraints that limit the flexibility the sandbox can offer by means of regulatory, technical, and real-world experimentation [46]. For instance, in Finland, from the point of view of the constitution, the purpose of the trial section must be socially acceptable to deviate from the principle of equality—for instance, business and innovation policy goals. Regarding exceptions, different types can be granted through trial clauses [42] or specific national laws published specifically for installing regulatory sandboxes [57]. For example, there can be easements to permit procedures. Still, because of the provisions in the AI Act, there is no foreseeable derogation (or exemption from the rule of law), because it would imply amendments that require action by EU-level regulators (European Commission within the European Union or the not yet instated European Board of Artificial Intelligence).

Nonetheless, sandbox experimentation can enable regulators to work with public sector organisations to ensure appropriate protection safeguards are built into the service before being widely deployed [46]. Mainly because in the EC AIA proposal, "significant risks to health and safety and fundamental rights identified during the development and testing of such system shall result in immediate" and "adequate mitigation, otherwise suspension until mitigation takes place or otherwise terminate it". Finally, in the EP AIA draft, "competent authorities shall have the power to temporarily or permanently suspend the testing process, or participation in the sandbox if no effective mitigation is possible and inform the AI office of such decision". Nonetheless, for the benefit of the provider which is participating in the regulatory sandbox to test their innovation the CE AI Act approach foresees "no fine

for infringement of the AI Act, if participants respect the sandbox plan, terms and conditions for participation, and there is good faith".

5.3 Design and Implementation

The third phase of the AI regulatory sandbox is the design and implementation. It is the practical experimentation with multiple stakeholders in the different stages of the AI lifecycle and within real-world conditions with a limited number of real end-users. In the EC AIA proposal, the AI regulatory sandbox is envisioned as a "controlled environment that facilitates the development, testing and validation of innovative AI systems''. The testing takes place "at any time before the placing on the market or putting into service of the AI system on their own or in partnership with one or more prospective users" [55]. Within our proposed framework, because the goal is to determine the usefulness of MLOps in compliance efforts, it would require this architecture to be part of the AI system design.

In terms of real-world testing, the CE AIA approach provides guidance on the details by establishing some of the general conditions as "drawn up a real-world testing plan and submitted to the market surveillance authorities of the member state where the testing is supposed to occur" and that "there is no objection by competent authorities within 30 days after its submission". For the provider, that it "is not in the area of law enforcement, migration, asylum and border control management and has registered the testing in the real-world conditions in the EU database", "is located in EU or has appointed a legal representative for the purpose of testing in real-world conditions which is established in the Union". For end-user/citizens, it should be ensured that "persons belonging to vulnerable groups due to their age, physical or mental disability are appropriately protected", "when testing with 1 or more prospective users, the participants shall be informed of all aspects of the testing that are relevant for the decision to participate and given the relevant instructions on how to use the AI system provider and user(s) shall conclude an agreement specifying their roles and responsibilities with a view to ensuring compliance with the provisions for testing in real-world conditions under this and other relevant regulation", "subjects of the testing in real-world conditions have given informed consent […], or in the case of law enforcement where the seeking of informed consent would prevent the AI system from being tested, the testing itself and the outcome of the testing in the real-world conditions shall not have a negative effect on the subject". For the testing conditions, that the "real-world conditions is effectively overseen by the provider and user(s) with persons who are suitable qualified in the relevant field and have the necessary capacity, training and authority to perform their tasks" and "the predictions, recommendations or decisions of the AI system can be effectively reversed and disregarded." Furthermore, "any serious incident identified in the course of the testing in real-world conditions shall be reported to the national market surveillance authority" and "provider shall establish a procedure for the prompt recall of the AI system upon such termination of the testing in real-world conditions" which we have depicted in Fig. 1 as retirement.

During the third phase, the compliance measures' experimentation can include specifying metrics for evaluation that can be participatory agreed upon (e.g., data quality of test data, reliability) and regularly used during the development as negative trends (e.g., not accurate predictions) can be detected and corrected at an early stage and addressed by the feedback loop [35]. Similarly, quality management could be achieved using pull requests as the basis for reviews and using model cards metadata as a tool for documenting not only the model but also the regulatory-specific activities performed during the model's development, such as dataset justification or performance evaluation or pre-market risk management activities [58]. If machine-readable, the model card metadata can be used in pipelines to automatically generate additional documents intended for end-users (e.g., the model card) and regulatory authorities (e.g., clinical validation report). Finally, monitoring and maintenance activities identifying deviations from the expected model behaviour are placed, captured as bug reports or feedback, and fed into the supervision authority for evaluation. This can help with issues around the explainability of models as it has been suggested [59] that "it's everything that happens around that decision they [experts using AI-based systems] need help with". Thus, because there is a need for 'more discussions about the output, rather than how you get there," monitoring, tracking, and documentation can help understand and contextualise the surrounding environment of an AI-based system's output.

Meanwhile, monitoring can help identify errors at an early stage. Using "triggering flags" implemented during the development helps determine when decisions need to be made before there is an improvement to the model. These "triggering flags" can be based on sandbox plan issues that were agreed upon as core to the experimentation (e.g., integration of human-on-the-loop under article 14 of the AI Act or need for risk assessment under article 9). Regarding the development of ML models, the metric figures collected via the interface can be used as the data basis for a decision. The basis for the decision could be a manual check which can base its review on the reports from the previous expansion stage. Based on a formalised evaluation of the key metrics (e.g., by checking for threshold value violations), it may also be possible to introduce a release decision that can be automated. In the deployment stage, the monitoring infrastructure that can detect deviations in the average accuracy and confidence of a deployed model can lead to the discovery of new input data that may relate to model drift or changes in the underlying relationships between input and output data, that may reveal the possibility for concept drift. Finally, the ML model corrective activities are part of the feedback loop that connects the monitoring stage to the building stage [35].

Based on the goals of a regulatory sandbox, core stakeholders need to ensure the utilisation of information by other stakeholders. For example, annual reports to the Board and Commission on the results (e.g., good practices, lessons learnt, and recommendations on their set-up) of the experimentation on the application of the AI Act and other Union legislation supervised within the sandbox [15]. The EC AIA proposal envisions a report "of the activities successfully carried out in the sandbox, results and learning outcomes," which could be available online following the steps of the Norwegian Regulatory sandbox. Furthermore, this report could be considered in the context of conformity assessment procedures or market surveillance checks.

In general, the potential outcomes of experimentation using AI regulatory sandboxes in the public sector include (1) sharing information as a way to ensure regulators can draw lessons from the sandbox, (2) developing more appropriate policies to foster innovation in the AI ecosystem, (3) facilitating the adoption of AI-based services in the public sector, (4) allowing regulators to gather evidence of potential needs for changes in existing regulatory frameworks, and (5) handling liability for negligence by allocating responsibility and providing practical redress via changes to AI models through MLOps and accountability for the responsible stakeholders.

6 Conclusion

The public sector faces distinct challenges incorporating AI-based systems to improve its services, including developing suitable technological tools for a multi-stakeholder regulated environment. We believe that AI regulatory sandboxes provide a space for limited experimentation beyond legal compliance and allocation of responsibility. In the case of "high-risk" AI systems, AI regulatory sandboxes and MLOps offer a potentially viable approach in an integrated framework that supports continuous experimentation and learning across the AI lifecycle in conjunction with multiple stakeholders for both technical validation and regulatory compliance. They support social acceptance of such AI-based services, greater public awareness of their implications, and better understanding and adoption in the public sector in high-risk scenarios.

This paper aims to provide a valuable framework for AI regulatory sandboxes which implement MLOps functionalities in their structure. There is also the potential benefit of experimenting with the use of MLOps as a tool for providing redress whenever there has been a mistake/faulty implementation through the lifecycle of an AI system. As MLOps enable monitoring, documentation, version keeping and retraining, it could allow for a correction within the model, which addresses issues raised by end-user. Those impacted by an AI system's output, or a regulator could, in a practical way, correct or request for an improvement in the model that prevents it from replicating such an error for other end-users and going beyond mere economic compensation or court proceedings. Thus, it can allow for a bottom–up approach to lawful AI systems.

MLOps can facilitate collaboration between different stakeholders via continuous monitoring, as once developed, the behaviour of an AI system has to remain as expected; otherwise, deficiencies such as drift would need to be addressed as early as possible [38]. In such regards, aspects of MLOps like continuous integration and deployment can help add new features that comply with regulatory body requests to a deployed model more rapidly. With proper regulatory oversight in a controlled sandbox environment, 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 the piloting of new features, models, and datasets while monitoring its technical and regulatory compliance, allowing both developers and regulators to continuously learn from iterative experimentation, mainly because in a more agile process like MLOps, where changes can be

implemented on a weekly or daily basis, possibly even independently of user intervention, effective auditing must be based on different principles.

In further research, we plan to validate the proposed framework in pilot tests of AI-enabled services developed by public sector organisations. Ongoing experimentation and research must be conducted with multiple stakeholders to examine how to support responsible AI innovation based on emerging regulatory compliance regimes, such as the AI Act. As a warning, there is the danger of creating misconceptions that once an AI system has been tested in a sandbox regulatory environment, it is given the stamp of approval to be placed in the market or put into service. It must be acknowledged that an AI system can still fail or produce harmful outcomes after general deployment. Thus, future implementations should clearly state that AI systems which have undergone experimental regulatory compliance could still produce unforeseen liability risks or evolve into high-risk AI through unanticipated applications. AI regulatory sandboxes offer only a limited timeframe to determine AI innovations' regulatory compliance measures before market placement or deploying them in a broader societal context. Nonetheless, experimental instruments, despite their shortcomings, contribute to the development of evidence-based law-making and the continuous reassessment of regulation [37], a more proactive approach to the interaction between law and technology. This paper proposes that AI regulatory sandboxes, in conjunction with integrated MLOps processes and practices, can offer crucial mechanisms for regulatory experimentation and technological innovation with multiple stakeholders across the lifecycle of AI-enabled services in the public sector.

Acknowledgements This work is undertaken as part of the research project Civic Agency in AI? Examining the AI Act and Democratizing Algorithmic Services in the Public Sector (CAAI) in the Critical AI and Crisis Interrogatives (CRAI-CIS) research group, the Department of Computer Science, Aalto University. The project is supported by research grants awarded by the Kone Foundation (2022-2025) and Research Council of Finland (2023-2027). We would like to extend our sincere thanks to the practitioners who volunteered to be interviewed for this research for their time, openness, and commitment. We also acknowledge support of our colleagues who have contributed to this research project and provided helpful feedback on this paper, particularly Nuuti Kytö from the City of Helsinki, and Kaisla Kajava in the CRAI-CIS Research Group.

Funding Open Access funding provided by Aalto University.

Declaration

Conflict of interest On behalf of the authors, the corresponding author states that there is no conflict of interest.

References

1. Floridi, L. (2020). Artificial intelligence as a public service: learning from Amsterdam and Helsinki. *Philosophy and Technology*, *33*(4), 541–546. https://doi.org/10.1007/s13347-020-00434-3.
2. Ruckenstein, M., Lomborg, S., & Hansen, S. S. (2020). Re-humanising automated decision-making. Workshop report from the ADM: Nordic Perspectives research network.
3. Haataja, M., van de Fliert, L., & Rautio, P. (2020). Public AI Registers. Realising AI transparency and civic participation in government use of AI. Whitepaper.
4. AI High Leven Expert Group. (2019). Ethics guidelines for trustworthy AI | Shaping Europe's digital future. Retrieved March 3, 2023, from https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.
5. European Commission. Joint Research Centre. (2022). AI Watch, road to the adoption of Artificial Intelligence by the public sector: A handbook for policymakers, public administrations and relevant stakeholders. Publications Office.
6. European Commission. (2021). Leveraging the power of location information and technologies to improve public services at the local level: State of the art report. Publications Office EU Joint Research Centre. https://doi.org/10.2760/158709.
7. Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., Myers West, S., Richardson, R., Schultz, J., & Schwartz, O. (2018). AI now report 2018. The AI Now Institute.
8. Chiusi, F., Fischer, S., Kayser-Bril, N., & Spielkamp, M. (2020). Automating society 2020. Algorithm Watch.
9. Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial Intelligence and the Public Sector—Applications and Challenges. *International Journal of Public Administration, 42*(7), 596–615. https://doi.org/10.1080/01900692.2018.1498103.
10. Androutsopoulou, A., Karacapilidis, N., Loukis, E., & Charalabidis, Y. (2019). Transforming the communication between citizens and government through AI-guided chatbots. *Government Information Quarterly*, *36*(2), 358–367. https://doi.org/10.1016/j.giq.2018.10.001.
11. D'Ignazio, C., & Klein, L. (2021). Data feminism (p. 328). MIT Press.
12. Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2023). Machine Bias. ProPublica. Retrieved March 3, 2023. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.
13. Parr, T. (2022). An audit of algorithms Nine algorithms used by the Dutch government. Netherlands Court of Audit.
14. Nikidehaghani, M., Andrew, J., & Cortese, C. (2022). Algorithmic accountability: Robodebt and the making of welfare cheats. *Accounting Auditing and Accountability Journal*, *36*(2), 677–711. https://doi.org/10.1108/AAAJ-02-2022-5666.
15. European Commission. (2021). Proposal for a regulation of the European parliament and of the council. Laying down Harmonised rules on artificial intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. No. COM (2021) 206 final (2021). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206.
16. OECD. (2021). Artificial intelligence: regulation can support innovation. Organisation for economic co-operation and development. https://www.oecd-ilibrary.org/science-and-technology/artificial-intelligence-regulation-can-support-innovation_f7fe0e1d-en.
17. Sawhney, N. (2022). Contestations in urban mobility: rights, risks, and responsibilities for urban AI. *AI and Society*. https://doi.org/10.1007/s00146-022-01502-2.
18. European Parliament. (2023). Draft compromise amendments on the draft report. Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206—C9 0146/2021—2021/0106(COD)). https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf.
19. European Commission. (2021). 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.
20. Dignum, V. (2022). Responsible artificial intelligence—from principles to practice. arXiv. https://doi.org/10.48550/arXiv.2205.10785.
21. European Parliament Research Service (EPRS). (2020). Artificial intelligence: From ethics to policy (Study Panel for the Future of Science and Technology PE 641.507).

22. Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S. J., O'Brien, D., Shieber, S., Waldo, J., Weinberger, D., & Wood, A. (2017). Accountability of AI under the law: the role of explanation. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3064761.
23. Kytö, N. (Forthcoming). Helsinki ML Ops. Machine Learning for City Services. Forthcoming White Paper, City of Helsinki.
24. Ojanen, A., Björk, A., Mikkonen, J., & Helsinki, D. (2022). Promoting equality in the use of Artificial Intelligence—an assessment framework for non-discriminatory AI. Policy Brief 2022:25. https://demoshelsinki.fi/wp-content/uploads/2022/08/promoting-equality-AI-2022.pdf.
25. Veale, M., & Borgesius, F. Z. (2021). Demystifying the draft EU artificial intelligence Act. *Computer Law Review International, 22*(4), 97–112. https://doi.org/10.9785/cri-2021-220402.
26. Madiega, T., & Van De Pol, A. L. (2022). Artificial intelligence act and regulatory sandboxes (Briefing PE 733.544). European Parliament.
27. Työ- ja elinkeinoministeriön julkaisuja. (2022). Innovaatiomyönteisen sääntelyn käytännöt kasvualoilla. Retrieved May 26, 2023, from https://tem.fi/innovaatiomyonteinen-saantely.
28. De Silva, D., & Alahakoon, D. (2022). An artificial intelligence life cycle: From conception to production. *Patterns*, *3*(6), 100489. https://doi.org/10.1016/j.patter.2022.100489.
29. Visengeriyeva, L., Kammer, A., Bär, I., Kniesz, A., Plöd, M., Machine learning operations. Retrieved May 26, 2023, from https://ml-ops.org/.
30. Ranawana, R., & Karunananda, A. S. (2021). An Agile Software Development Life Cycle Model for Machine Learning Application Development. 2021 5th SLAAI International Conference on Artificial Intelligence (SLAAI-ICAI), 1–6. https://doi.org/10.1109/SLAAI-ICAI54477.2021.9664736.
31. Bettinger, K., Ziskind, J., & Lähteenoja, V. (2021). Empowered data societies: A human-centric approach to data relationships. White Paper. World Economic Forum. https://www.weforum.org/whitepapers/empowered-data-societies-a-human-centric-approach-to-data-relationships/.
32. Amnesty International. (2021). Dutch childcare benefit scandal an urgent wake-up call to ban racist algorithms. Amnesty International. Retrieved March 3, 2023, from https://www.amnesty.org/en/latest/news/2021/10/xenophobic-machines-dutch-child-benefit-scandal/.
33. Engler, A., & Renda, A. (2022). Reconciling the AI Value Chain with the EU's Artificial Intelligence Act. Centre for European Policy Studies (CEPS). https://www.ceps.eu/ceps-publications/reconciling-the-ai-value-chain-with-the-eus-artificial-intelligence-act/.
34. 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*, *11*(9), 2149–2170. https://doi.org/10.14807/ijmp.v11i9.1400.
35. Stirbu, V., Granlund, T., & Mikkonen, T. (2022). Continuous design control for machine learning in certified medical systems. *Software Quality Journal*. https://doi.org/10.1007/s11219-022-09601-5.
36. Pechtor, V., & Basl, J. (2022). Analysis of suitable frameworks for artificial intelligence adoption in the public sector. Digitalization of Society, business and management in a pandemic. 30th Interdisciplinary Information Management Talk. https://doi.org/10.35011/IDIMT-2022-67.
37. Ranchordas, S. (2021). Experimental regulations for AI: sandboxes for morals and mores (SSRN Scholarly Paper No. 3839744). *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3839744.
38. Leslie, D., Burr, C., Aitken, M., Katell, M., Briggs, M., & Rincon, C. (2021). Human rights, democracy, and the rule of law assurance framework for AI systems: A proposal. Alan Turing Institute. *SSRN Journal*. https://doi.org/10.2139/ssrn.4027875.
39. Armstrong, H., Gorst, C., & Rae, J. (2019). Renewing regulation. 'Anticipatory regulation' in an age of disruption. Nesta. Retrieved May 26, 2023, from https://www.nesta.org.uk/report/renewing-regulation-anticipatory-regulation-in-an-age-of-disruption/.
40. Schaefer, C., Lemmer, K., Kret, K. S., Ylinen, M., Mikalef, P., & Niehaves, B. (2021). Truth or dare?—How can we influence the adoption of artificial intelligence in municipalities?. In *Proceedings of the Annual Hawaii International Conference on System Sciences* (pp. 2347–2356). https://doi.org/10.24251/hicss.2021.286.
41. Council of the European Union. (2020). Regulatory sandboxes and experimentation clauses as tools for better regulation: Council adopts conclusions. https://www.consilium.europa.eu/en/press/press-releases/2020/11/16/regulatory-sandboxes-and-experimentation-clauses-as-tools-for-better-regulation-council-adopts-conclusions/.
42. Federal Ministry for Economic Affairs and Energy (BMWi). (2019). Making space for innovation. The handbook for regulatory sandboxes. Retrieved May 26, 2023, from https://vdocuments.net/making-space-for-innovation-bmwi.html.

43. Bauknecht, D., Heyen, D., Gailhofer, P., Bizer, K., Feser, D., Führ, M., Winkler-Portmann, S., Bischoff, T., & Proeger, T. (2021). How to design and evaluate a Regulatory Experiment? A Guide for Public Administrations. Retrieved May 26, 2023, from https://reragi.files.wordpress.com/2019/04/regulatory_experiments-guide_for_public_administrations.pdf.
44. Opinion of Advocate General Maduro in Arcelor Atlantique case (C-127/07, EU:C:2008:728). (2017). Retrieved May 26, 2023, from https://curia.europa.eu/juris/document/document.jsf?docid=188755&doclang=EN.
45. Financial Conduct Authority (FCA). (2015). Regulatory sandbox.
46. Truby, J., Brown, R. D., Ibrahim, I. A., & Parellada, O. C. (2022). A Sandbox approach to regulating high-risk artificial intelligence applications. *European Journal of Risk Regulation, 13*(2), 270–294. https://doi.org/10.1017/err.2021.52.
47. Nuno, N., De Andrade, G., & Zarra, A. (2022). Artificial Intelligence Act: A Policy Prototyping Experiment. Operationalizing the Requirements for AI Systems—Part I. Open Loop. Retrieved 26, 2013, from https://openloop.org/news/open-loop-report-artificial-intelligence-act-a-policy-prototyping-experiment/.
48. Omarova, S. T. (2020). Technology v technocracy: Fintech as a regulatory challenge. *Journal of Financial Regulation, 6*(1), 75–124. https://doi.org/10.1093/jfr/fjaa004.
49. UNSGSA FinTech Working Group and CCAF. (2019). Early Lessons on Regulatory Innovations to Enable Inclusive FinTech: Innovation Offices, Regulatory Sandboxes, and RegTech. Retrieved May 26, 2023, from https://www.unsgsa.org/sites/default/files/resources-files/2020-09/UNSGSA_Report_2019_Final-compressed.pdf.
50. Hellsten, J. (2021). A place of growth. Helsinki City Strategy 2021–2025. City of Helsinki. https://www.hel.fi/static/kanslia/Julkaisut/2021/helsinki-city-strategy-2021-2025.pdf.
51. Ringe, W. G., & Ruof, C. (2020). Regulating fintech in the EU: The case for a guided sandbox. *European Journal of Risk Regulation, 11*(3), 604–629. https://doi.org/10.1017/err.2020.8.
52. GitHub, continuous integration and continuous delivery (CI/CD) Fundamentals. GitHub Resources. Retrieved May 26, 2023, from https://resources.github.com/ci-cd/.
53. Souris, H., Großmann, J., Himberg, J., Heikki, I., & Knoblauch, D. (2023). Initial MLOps methodology and the architecture of the IML4E framework. Silo AI. Retrieved May 26, 2023, from https://iml4e.org/85e3ef5b3db3736f.
54. Granlund, T., Stirbu, V., & Mikkonen, T. (2021). Towards Regulatory-compliant MLOps: Oravizio's journey from a machine learning experiment to a deployed certified medical product. *SN Computer Science, 2*(5), 342. https://doi.org/10.1007/s42979-021-00726-1.
55. Council of the European Union. (2022). Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts—General approach.
56. Mehr, H. (2017). Artificial Intelligence for Citizen Services and Government. Harvard Kennedy School. ASH Center for Democratic Governance and Innovation. Retrieved May 26, 2023, from https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf.
57. Ley 7/2020, de 13 de noviembre, para la transformación digital del sistema financiero. Jefatura del Estado. BOE-A-2020-14205. Retrieved May 26, 2023, from https://www.boe.es/eli/es/l/2020/11/13/7.
58. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 220–229). https://doi.org/10.1145/3287560.3287596.
59. Liao, Q. V., Gruen, D., & Miller, S. (2020). Questioning the AI: Informing design practices for explainable AI user experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–15). https://doi.org/10.1145/3313831.3376590.