AI Regulatory Sandboxes between the AI Act and the GDPR: the role of Data Protection as a Corporate Social Responsibility

Davide Baldini[1,*,†] and Kate Francis[2,*,†]

[1] *University of Florence, Via delle Pandette, 32, 50127, Florence, Italy*
[2] *Maastricht University, Bouillonstraat 3, 6211 LH Maastricht, Netherlands*

Abstract

This paper investigates the potential for regulatory sandboxes, a new and innovative regulatory instrument, to improve the cybersecurity posture of high-risk AI systems. Firstly, the paper introduces AI regulatory sandboxes and their relevance under both the AI Act and the GDPR. Attention is paid to the overlapping cybersecurity requirements derived from both pieces of legislation. The paper then outlines two emerging challenges of AI cybersecurity. The first, factual challenge, relates to the still under-developed state-of-the-art of AI cybersecurity, while the second legal challenge relates to the overlapping and uncoordinated cybersecurity requirements for high-risk AI systems stemming from both the AI Act and GDPR. The paper argues that AI regulatory sandboxes are well-suited to address both challenges which, in turn, will likely promote the uptake of AI regulatory sandboxes. Subsequently, it is argued that this novel legal instrument aligns well with emerging trends in the field of data protection, including Data Protection as Corporate Social Responsibility and Cybersecurity by Design. Taking stock from this ethical dimension, the many ethical risks connected with the uptake of AI regulatory sandboxes are assessed. It is finally suggested that the ethical and corporate social responsibility dimension may provide a potential solution to the many risks and pitfalls of regulatory sandboxes, although further research is needed on the topic.

Keywords

Regulatory sandboxes, AI Act, GDPR, Cybersecurity, Data Protection as a Corporate Social Responsibility, Ethics

1. Introduction

Regulatory sandboxes can be described as controlled spaces where authorities engage firms to test innovative products or services that challenge existing legal frameworks, for a limited amount of time (OECD, 2023). While many different types of regulatory sandboxes exist across several jurisdictions, some common characteristics are typically present, namely: their temporary nature, the involvement of both regulators and firms, the waiving of existing legal provisions (and related liability), the provision of tailored legal support for a specific project from the regulator, as well as the possibility for the latter to acquire and use

ITASEC 2024: The Italian Conference on CyberSecurity, April 08–11, 2024, Salerno, Italy
[1*] Corresponding author.
[†] These authors contributed equally. Authors are PhD candidates at their respective universities
davide.baldini@unifi.it (D. Baldini); k.francis@maastrichtuniversity.nl (K. Francis)

technical and market information, which can be used to assess whether specific legal frameworks are fit-for-purpose or need to be adapted (OECD, 2023).

Given their flexible and dynamic nature, regulatory sandboxes are increasingly being regarded as a promising instrument, capable of overcoming many of the traditional shortcomings inherent to the regulation of new technologies (OECD, 2023; Bagni, 2023). As a result of this, over the last few years, regulatory sandboxes are increasingly leveraged by governments and organizations as a useful tool for regulating new technologies, especially in the context of FinTech (Bagni, 2023) and, more generally, AI-powered technologies: according to the OECD, as of May 2023 there were "about one hundred sandbox initiatives around the world (...), including fintech and privacy sandboxes" (OECD, 2023). Regulatory sandboxes also come with many risks including abuse and misuse (Junklewitz et al., 2023) and therefore necessitate an ethical approach by all parties involved.

The first objective of this paper is to analyze this new innovative regulatory instrument and its potential application in the context of the cybersecurity of Artificial Intelligence systems, with a primary focus on both the General Data Protection Regulation ("GDPR") and the draft AI Act. While the AI Act expressly envisages the creation of regulatory sandboxes within its Title V ("measures in support of innovation"), the GDPR is not concerned with establishing similar tools for testing data protection compliance. However, and despite the lack of reference to sandboxes within the GDPR, the latter requirements – including the cybersecurity requirements outlined in its Article 32 – will be relevant for the operation of most AI regulatory sandboxes, given that high-risk AI systems typically leverage personal data in the context of their development and operation and, as a result, the GDPR scope of application is triggered.

This intersection between the rules set forth by the AI Act and GDPR as regards the establishment and functioning of regulatory sandboxes is already acknowledged and partially addressed by Article 54 of the AI Act, although many open questions remain on how the two legal instruments will interact in the context of Regulatory Sandboxes.

Building upon said analysis, this paper explores the potential role of regulatory sandboxes in the context of the nascent field of Data Protection as Corporate Social Responsibility ("DPCSR"), with a specific focus on its potential to foster the consideration of ethical principles when new solutions are experimented with (Balboni and Francis, 2023). In doing so, it also aims to point towards the need for careful consideration of ethical implications in the context of the establishment and functioning of Regulatory Sandboxes and opens the doors for future research on this topic.

2. Cybersecurity of high-risk AI systems: Factual and legal challenges

The draft AI Act is concerned with AI systems, defined within its Article 3.1.1 as machine-based systems that are "designed to operate with varying levels of

autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions, that influence physical or virtual environments".

In light of this broad definition, an AI system is to be regarded as a computer system that includes one or more AI models, along with other non-AI components (Junklewitz et al., 2023). Take, for instance, a chatbot such as ChatGPT: the underlying AI component is the pre-trained Large Language Model (in the example, GPT 3.5 or GPT 4.0), which is indeed an essential element of the system, but by no means the only one; in order for an AI system such as ChatGPT to function, a number of other non AI-related components must be present, such as the cloud infrastructure, pre-processing software, backup systems, and so on (Junklewitz, et al., 2023). Even when taking into account, the AI model as an integral component of the AI System, the latter remains, at its core, essentially a software. Therefore, from a cybersecurity perspective, and as with any other computer program, AI systems share existing security risks of traditional software, while also adding new risks specifically related to the AI model component of the AI systems. Noteworthy examples are data poisoning, adversarial attacks and model poisoning (ENISA 2023).

2.1. The factual challenge: A state-of-the-art undergoing development

Article 15 of the AI Act addresses the cybersecurity requirements of AI systems, within the context of the broader conformity assessment procedure foreseen in Article 9 of the AI Act. By applying to the AI system as a whole, and not only to the AI model, this norm requires that cybersecurity be addressed vis-à-vis all components of the AI systems. However, while existing cybersecurity practices of proven effectiveness may be successfully applied to the non-AI components of AI systems, specific cybersecurity measures must be implemented in case of the AI model, as recalled within the last paragraph of Article 15 of the AI Act.

Current cybersecurity practices, while effective for "traditional" computer systems, are not suited to address this new spectrum of AI-specific cybersecurity risks (ENISA, 2023). The nascent field of AI cybersecurity focuses on researching and mitigating these AI-specific vulnerabilities, by developing new practices and procedures tailored to secure AI models (Junklewitz et al., 2023). This effort includes the creation of AI cybersecurity risk management tools, security controls, metrics, and mitigation measures designed specifically to curtail risks such as data poisoning, model poisoning, adversarial attacks, and so on.

In this respect, Article 15 of the AI Act essentially requires that the cybersecurity measures aimed at of high-risk AI systems be appropriate to the relevant risks and circumstances in which the AI system operates. In practice, also in the light of recital 51, high-risk AI systems will have to be designed and developed following the principle of security by design and by default, by means of implementing state- of-the-art measures according to the relevant market segment or scope of application, so as to achieve an appropriate level of cybersecurity, to be maintained consistently

throughout the system lifecycle. Compliance with this obligation will arguably require the provider to carry out a cybersecurity risk assessment, to be performed in the context of the broader risk management system outlined in Article 9 of the AI Act. Given that, as observed above, the AI Act is concerned with AI systems (and not only with the AI model, as a sub-component of the system), the assessment should concern both general and AI-specific cybersecurity vulnerabilities, as well as mitigation measures, in light of the architecture and purpose of the system.

The effective mitigation of AI-specific vulnerabilities, however, presents multiple challenges. Given the recent emergence of the technology and its fast-paced and dynamic nature, the state-of-the-art of the AI cybersecurity field is still considered to be somewhat lagging behind the state-of-the-art of the technology (Junklewitz et al., 2023). In some cases, it may even be currently impossible, for applications using the most innovative AI techniques, to comply with the cybersecurity requirements of the AI Act (Junklewitz et al., 2023). While the level of sophistication of AI technologies continuously increases, as made evident by the multiple breakthroughs that have recently taken place in the field of generative AI, this problem might be here to stay, with state-of-the-art cybersecurity practices struggling to catch-up with the most recent technological advances.

This state of affairs presents the question of how innovative AI systems can continue to be developed and marketed in the EU, while being at the same time in compliance with the stringent cybersecurity requirements of the AI Act. As shall be seen in Section 3, AI regulatory sandboxes are well-equipped to assist providers in establishing an adequate level of cybersecurity posture for their innovative high-risk AI systems.

2.2. The legal challenge: Multiple cybersecurity requirements in search of coordination

Despite the AI Act ambition to provide a clear and exhaustive, it should be underlined that Article 15 of the AI Act does not exhaust the cybersecurity obligations for high-risk AI systems. Most AI systems typically leverage personal data both in the context of their development, deployment and operation, which means that EU data protection law will continue to apply to any personal data processing activity carried out by the AI system. In this respect, Article 2 of the AI Act recognizes that the AI Act "shall not affect" the applicability of EU data protection law, including the GDPR. As a result of this, the cybersecurity requirements of high-risk AI systems will be regulated not only by the aforementioned Article 15 AI Act, but also by Article 32 GDPR, thereby including the body of regulatory guidance and Data Protection Authorities' (DPAs) case-law that has developed over the years concerning this provision.

The two provisions not only have different wordings and partially different scope but are also subject to different enforcement authorities. This means that both GDPR-established data protection authorities and AI Act-established market surveillance authorities can claim jurisdiction over the cybersecurity aspects of high-risk AI systems. It can be expected that this double exposure to enforcement

actions by different authorities with jurisdiction over different legal instruments will create regulatory uncertainty, presenting providers of high-risk AI systems with, at best, unclear and, at worst, conflicting requirements. This risk seems especially concrete when considering the eagerness already shown by several data protection authorities in enforcing data protection rules vis-à-vis providers and deployers of AI systems as seen – most notably – in the Italian Garante's enforcement actions against OpenAI between 2023 and 2024. Aside from the Garante, several other EU Member States data protection authorities are de facto already acting as AI regulators, especially in the context of generative AI (Zanfir-Fortuna, 2023).

This overlap between the AI Act and the GDPR cybersecurity requirements for high-risk AI systems is not addressed by the two bodies of legislation, nor – to date – has any regulatory guidance been issued on the topic. Thus, the question remains on how the two overlapping provisions can be coordinated in a way that ensures legal certainty for providers of high-risk AI systems that process personal data. However, as shall be seen in the following section, the AI regulatory sandboxes envisaged by the AI Act are a promising tool to curtail this issue.

3. AI Regulatory Sandboxes as a potential solution

Sandboxes are foreseen in Section V of the draft AI Act, titled "Measures in support of innovation". In particular, Article 53 mandates the establishment of one or more regulatory sandboxes at national level, by each Member State or jointly by more than one, while also envisaging the possibility to establish regional or local sandboxes. For the purposes of the AI Act, an AI regulatory sandbox is described in Article 52(1d.) as "a controlled environment that fosters innovation and facilitates the development, training, 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 sandbox plan agreed between the prospective providers and the competent authority", further adding that "(...) regulatory sandboxes may include testing in real world conditions supervised in the sandbox".

Enrollment in an AI regulatory sandbox grants the participating provider of high-risk AI systems with many benefits. Most notably, competent authorities involved in the sandbox are required to provide to the participating organizations "as appropriate, guidance, supervision and support within the sandbox" (Art. 53.1d), as well "guidance on regulatory expectations and how to fulfil the requirements and obligations set out in this Regulation" (Art. 53.1f). While doing so, the competent authorities may not impose administrative fines to participating providers who have followed their guidance in good faith, as well as the specific terms and plan for participation (Art. 53.4).

From their part, regulators who attend the sandbox are enabled to acquire technical insights and information on new AI technologies and/or novel technological applications of AI which are being developed by providers, way before they are marketed; as such, AI sandboxes establish a new venue of evidence-based,

ex ante regulatory learning for supervisory authorities (Art. 53.1g.(d)), thanks to which supervisory authorities will be able to better calibrate and fine-tune their enforcement activities, by being able to stay up-to-date with, and gain useful insights into, new and emerging technological trends, a phenomenon known as "regulatory learning" (Ranchordas & Vinci, 2024).

Providers who attend the sandbox are therefore able to obtain tailored compliance advice on their AI systems, directly from the same supervisory authorities who are tasked with enforcing the relevant rules. From a cybersecurity perspective, such advice will concern, inter alia, the specific measures which the regulators deem adequate to address the vulnerabilities presented by the AI system which is undergoing the sandbox process.

It is safe to assume that this type of tailored regulatory advice will prove especially useful for providers of innovative high-risk AI systems which, for the reasons outlined in the previous section, will face a limited and not fully mature state of the art for AI-specific threats. This advantage could prove especially in the early period of the AI Act enforcement, when it can be expected that guidance issued by competent authorities – such as the AI office – will be scant. As a result, providers that manage to obtain direct advice from enforcement authorities will gain a compliance advantage over competitors, especially in the first period of applicability of AI Act.

In addition, for those high-risk AI systems that process personal data and are thus also regulated by the GDPR, Article 53, paragraph 2, of the AI Act, expressly requires that data protection authorities be "associated to the operation of the AI regulatory sandbox and involved in the supervision of those aspects to the extent of their respective tasks and powers". As a result, the competent data protection authority will be required to attend the AI regulatory sandbox and to provide coordinated guidance to the participating providers concerning, inter alia, regulatory expectations related to the cybersecurity posture of high-risk AI systems.

The possibility for providers of high-risk AI systems to attend a regulatory sandbox and thus benefit from coordinated guidance issued jointly by both the AI Act and GDPR competent authorities is particularly welcome given that, as observed in Section 2.2, Article 15 of the AI Act does not exhaust the cybersecurity obligations for high-risk AI systems, which must also achieve compliance with Article 32 GDPR, to the extent that they process personal data.

This arrangement will likely prove much practical value for AI providers as, in practice, high-risk systems typically leverage personal data both in the context of their development and operation, which in turn means that the GDPR routinely applies to the processing activities carried out by the AI system, even when they take place in the context of the sandbox.

As we have seen, AI regulatory sandboxes are a promising instrument, which is well-suited to, on the one hand, advance the state-of-the-art of AI cybersecurity while, on the other hand, forcing GDPR and AI Act enforcement authorities to work together and provide joint guidance on cybersecurity requirements. In light of this, it is reasonable to assume that cybersecurity will be an important driver to entice

providers of innovative high-risk AI systems to attend AI regulatory sandboxes. As a result, it can reasonably be expected that this regulatory instrument will see many applications, especially in the first period of the AI Act applicability, when publicly-available regulatory guidance on AI cybersecurity will be especially lacking.

A further element that may promote the uptake of AI regulatory sandboxes, as shall be seen in the following Section, is the fact that their characteristics align with emerging trends in the data protection landscape, that is, Data Protection as Corporate Social Responsibility and Cybersecurity by Design.

4. Sandboxes in the light of Data Protection as a Corporate Social Responsibility and Cybersecurity by Design

As anticipated above, regulatory sandboxes have two primary functions. The first is that of fostering both "the development and testing of innovations in a real-world environment" and the second is that of assisting regulators in "regulatory learning", defined as "the formulation of experimental legal regimes to guide and support businesses in their innovation activities under the supervision of a regulatory authority" (Madiega & Van De Pol, 2022, p. 2). The potential beneficiaries of regulatory sandboxes are multiple and include regulators, industry, consumers and generally, society. Despite offering many benefits, sandboxes also present a multitude of ethical and social risks which must be successfully and transparently mitigated for the benefits to be truly reaped.

This section provides an overview of some of the benefits and risks posed by regulatory sandboxes and provides initial insights as to how ethics and sustainability can contribute to such mitigation by framing cybersecurity and privacy under the umbrellas of Corporate Social Responsibility ("CSR") and Environmental, Social, and Governance ("ESG") via the Maastricht University Data Protection as a Corporate Social Responsibility Framework ("UM-DPCSR Framework") (Balboni & Francis, 2023).

Sandboxes are particularly attractive in the current regulatory scenario given the difficulty of developing future-proof legislation and the fast speed at which technology develops. The German Federal Ministry for Economic Affairs and Energy (2019) has confirmed the potential for regulatory sandboxes to manage "regulatory issues in the field of sustainability, the sharing economy and digital administration" (p. 13). This potential for sustainability to be incorporated into technologies thanks to their development in sandboxes suggests a connection with ESG and resonates especially in the area of AI, which is developing at an unprecedented pace.

For some time, the Norwegian Data Protection Authority has overseen a data protection regulatory sandbox focused on AI. The Norwegian DPA's sandbox specifically aims "to stimulate privacy- enhancing innovation and digitalization" (Datatilsynet, n.d.). The AI regulatory sandbox guides a select group of diverse companies operating in an array of sectors "in exchange for full openness about the assessments that are made" (Datatilsynet, n.d.). Coherent with the increasing necessity for data ethics to be considered in technology development, a primary

objective pursued is to "promote the development of innovative solutions that, from a data protection perspective, are both ethical and responsible" (Datatilsynet, n.d.).

Sandboxes have the potential to permit cybersecurity, privacy, data protection, fairness, user empowerment, and data ethics by design and sustainability considerations to be built into new technological innovations from their ideation thanks to oversight from authorities. By authorities ensuring that ethical aspects are taken into consideration by companies, sandboxes can thus contribute to lawful, sustainable and ethical processing (Balboni & Francis, 2023). Such a by-design approach is in line with both EU legislation and what is proposed in the UM-DPCSR Framework developed at the European Centre on Privacy and Cybersecurity - the world's first auditable framework comprised of five principles, 25 rules and 44 controls on ethical and socially responsible data processing.

The premise of the UM-DPCSR Framework is that in our data-driven world, there is a need to move towards an ethical approach that goes beyond the law when it comes to the processing of data and the development of new technologies (Balboni & Francis, 2023). Given the current enforcement scenario in the EU, and in light of the fact that not all legal activities may be ethical or lead to a benefit for society or individuals (beyond the mere benefit of economic growth) organizations, to be sustainable in the long-term, must start to consider data protection and cybersecurity as assets and moral imperatives as opposed to mere compliance obligations (Balboni & Francis, 2023).

The UM-DPCSR Framework adopts a soft-law approach that complements traditional enforcement of economic actors by independent administrative authorities. Legal compliance is furthermore a prerequisite to adhere to the Framework (Balboni & Francis, 2023). The UM-DPCSR Framework calls upon organizations to rethink the ways in which they approach privacy, cybersecurity, and in general, the use of data and the development of technologies. It requires that organizations act transparently and take all measures possible to protect fundamental rights and engage in practices which are not specifically coded in law, but which produce positive benefits for society. At the same time, it acknowledges the necessity of businesses to create economic profits and incentivizes compliance thanks to the potential for them to increase trust and demonstrate their accountability.

The Framework translates theoretical ethical principles into concrete and auditable actions which can be followed by organizations to engage in data processing that produces tangible benefits for society (Balboni & Francis, 2023). This includes, for example, cybersecurity by design. Because sandboxes may entail adaptions in the application of legislation, ethics and transparency are paramount to ensure ethical and cybersecure technologies that do not harm individuals. Due to the fact that ethical best practices are notoriously nebulous, the auditable UM-DPCSR Framework can act as a promising roadmap for organizations and authorities alike wishing to ensure sustainable, ethical practices as well as additional considerations which may not be specifically required by law but that may produce positive outcomes for society.

By participating in regulatory sandboxes and building ethics and compliance into innovations thanks to a close working relationship with the authorities and consideration of ethics (e.g., considering the 44 controls of the Framework), organizations can ensure that their solutions are socially responsible because they – by default – comply with the law and even go beyond it in the case that ethical principles (not codified in law) are adhered to.

5. The risks of AI Regulatory Sandboxes

Risks of abuse and misuse, and specifically the lowering of safeguards with the aim of attracting innovators, regulatory capture, the prioritization of innovation as opposed to safeguards, a potential "race to the bottom", regulatory fragmentation, market fragmentation, the slowing down of innovation, and the infliction of harm as a result of failure to comply with applicable rules and ethics best practices are all possible risks in the context of sandboxes (Ranchordas & Vinci, 2024, p. 107; Parenti 2020; Madiega & Van De Pol, 2022, p. 3). Within sandboxes, a diverse array of actors interact: "developers, regulators, experts, and consumers of innovative products" – leading to a "more consensual approach to defining the applicable rule" (Madiega & Van De Pol, 2022, p. 2). While such consensual rulemaking is attractive, it comes with significant risks. This section tackles a selection of risks identified in research on the topic.[2]

Within sandboxes, developers may be in the position to pressure or corrupt officials into prioritizing technological innovation over ethical or legal safeguards (Madiega & Van De Pol, 2022, p. 3). This would have a clear negative impact on consumers in the long run. Similarly, as Parenti (2020) suggests, a "race to the bottom", could lead to lowering regulatory safeguards in the aim of, for example, attracting income (think about Ireland and Luxembourg's enforcement of the GDPR which is notoriously critiqued). For this reason, clear rules of the game must be established, also to ensure accountability of the parties. Regulators must furthermore establish transparent objectives to inform the expectations of companies participating and, according to Parenti (2020), an internal review process to ensure effectiveness of the sandbox and decisions taken by the DPA is also recommendable (p. 9).

A lack of transparency is furthermore seen in the context of the relationship between the regulator and companies participating in sandboxes, something that Ranchordas and Vinci (2024) call a "double- edged sword" necessitating a good level of transparency facilitated through information sharing between both the regulator and participants, amongst participants themselves, and also in the form of reports evaluating participation in the sandbox (p. 110). Such transparency has the potential to more widely diffuse the benefits of sandboxes, making them more worthwhile and socially relevant. According to Ranchordas and Vinci (2024), failure to ensure transparency in sandboxes essentially "limit[s] the ability of stakeholders outside

[2] Due to word limits, this section merely aims to present a selection of risks which have been identified in the literature and does not aim to be complete.

the sandbox to scrutinize the equity of its measures, potential competitive advantages conferred to sandbox participants, and hold regulators accountable for agency drift" (p. 110). Transparency, in this sense can act as a further level of control to mitigate unlawful but also unethical and non-beneficial behavior within sandboxes.

Due to the fact that within sandboxes, there may be situations of regulatory uncertainty, a lack of adequate preexisting rules, and a desire to act outside the legal framework, regulatory sandboxes necessitate an ethical approach in the management of the relationship between the authority and the innovating company. But who decides what is ethical and what is not? (Undheim et al., 2023, p. 999). As Undheim et al. (2023) suggest, regulatory sandboxes may be an ideal place to "for exploring the boundaries of ethics, exploring hypothetical risk and uncertainties, or more fundamentally, fostering a moral imagination" (p. 999). While this may be true, such exploration would seem to easily lead to a situation of regulatory capture, suggesting that it is favorable to agree on ethical principles to be applied within sandboxes with a wider stakeholder group that does not have a vested interest in a particular outcome of regulation.[3]

Regulatory capture, the idea that regulators tend to "identify with the interest of the industry they are supposed to regulate… rather than the interests of its customers, or the general public" (Oxford reference, n.d.), is a serious risk in the context of regulatory sandboxes. There are various ways in which regulatory capture could unfold. It may be the case, for example, that companies purposefully attempt to influence regulators to act in their own interest, potentially at the expense of society (Ranchordas & Vinci, 2024, p. 110). It is thus clear that regulatory capture represents a serious risk with respect to the identify with the interest of the industry they independence of DPAs.

Article 52, paragraph 1, GDPR, requires that DPAs are completely independent in the performance of their tasks and the exercise of their powers and Article 52(2) GDPR requires that members of DPAs "remain free from external influence, whether direct or indirect, and shall neither seek nor take instructions from anybody". The necessity for supervisory authorities to maintain their independence and to avoid any kind of influence appears to be challenged in the context of regulatory sandboxes where due to the close working relationship between the regulator and the company in the context of the sandbox, this may be difficult. Such potential may furthermore be exacerbated in cases where transparency regarding decision-making is not ensured (Ranchordas & Vinci, 2024, pp. 135-136).

Sandboxes provide an opportunity for supervisory authorities to learn about technological advancements and trends within specific sectors and companies which they otherwise may not have had insights into. In a world where technologies can be used for good or for evil, it is opportune that authorities are in tune with

[3] The UM-DPCSR Framework, for example, identifies a series of "ethical" principles including but not limited to those identified by the High Level Expert Group on AI established by the European Commission.

relevant sector activities. Sandboxes in this sense also provide DPAs with technical knowledge that they otherwise may not have had access to due to limited budgets which are notoriously a cause of poor enforcement (Datatilsynet, n.d.). Precisely because sandboxes permit authorities to make use of relevant insights made during the sandbox to draft new guidance aimed, for example, to organizations developing and deploying AI systems – allowing also the authority to develop new competencies when it comes to AI, it is fundamental that such knowledge is shared with a wider public (Datatilsynet, n.d.). Such a public may be comprised of other authorities, other companies operating in the same sector as the organization within the sandbox, companies operating in other sectors, and the wider public (Parenti, 2020). Only where knowledge and insights are openly shared can their impact be amplified.

Lastly, there is a threat of risk-washing. Risk-washing could lead to authorities directly contributing to fundamental rights violations. Brown and Piroska (2022) define risk-washing as an "analogy with greenwashing, as a (...) regulatory institution's making products or processes of a company seem to involve less risk for stakeholders by engaging in activities that mimic in a superficial or narrow way genuine attempts to assess and reduce risk" (p. 20). To avoid risk-washing, it is necessary to "verify" the rules of engagement and ensure that they are carefully followed. The Norwegian method of involving both internal and external experts could be considered a best practice in this sense.[4]

In short, the risks posed by regulatory sandboxes are many, even if only a small number are discussed in this paper. Further research on the risks and possible mitigation measures is needed to ensure that the noble objective of the legislator in the AI Act is accomplished.

6. Conclusion: corporate social responsibility as the way forward

As we have seen, due to their characteristics, AI regulatory sandboxes seem particularly well-suited not only to advance the state-of-the-art of AI cybersecurity, but also to provide participating companies with coordinated guidance on regulatory expectations concerning the overlapping cybersecurity requirements stemming from the GDPR and the AI Act.

Furthermore, AI regulatory sandboxes seem to align pretty well with emerging trends in the data protection landscape, including data protection as corporate social responsibility and cybersecurity by design.

While still under-explored, these are positive elements of regulatory sandboxes that may contribute to the uptake of this new legal instrument. However, the diffusion of AI regulatory sandboxes is not without its issues.

[4] "The sandbox selection committee consists of an internal, interdisciplinary group that conducts interviews with all applicants. An external reference group, comprising members from Innovation Norway, the Norwegian Computing Centre, the Equality and Anti-Discrimination Ombud and Tekna, will assist in assessing the public benefit of the potential projects. The final selection of projects accepted into the sandbox will be made by the steering committee, made up by the Authority's management." (Makussen, 2023, p. 17).

Due to the fact that regulatory sandboxes may lead to deviations from the current legal framework, including from the GDPR and AI Act, harms may be inflicted upon the fundamental rights and freedoms of individuals.

The establishment and following of baseline ethical principles and red lines for both the participating regulators and for companies, going beyond the express legal requirements, is thus fundamental to ensure that the many risks connected to the uptake of AI regulatory sandboxes are properly addressed. In this sense, data-protection focused CSR frameworks such as the UM-DPCSR may act as a yardstick for supervisory authorities and companies to ensure that data processing activities carried out in the context of sandboxes are socially beneficial and ethical, in addition to being lawful. However, there is a significant need for further research on the topic of sandboxes, in particular for what regards ethical aspects, which are currently understudied (Ranchordas, 2021).

References

[1] OECD (2023). *Regulatory sandboxes in artificial intelligence*. OECD Digital Economy Papers, No. 356, OECD Publishing, DOI: 10.1787/8f80a0e6-en.

[2] Bagni, F. (2023). *La sandbox regolamentare e la sfida a tema cybersecurity: dall'Artificial Intelligence Act al Cyber Resilience Act*. Rivista Italiana di Informatica e Diritto, 5(2), 201-217, DOI: 10.32091/RIID0119.

[3] Junklewitz, H., Hamon, R., André, A., Evas, T., Soler Garrido, J. and Sanchez Martin, J.I. (2023). *Cybersecurity of Artificial Intelligence in the AI Act*, Publications Office of the European Union, Luxembourg, 2023, doi: 10.2760/271009.

[4] Balboni, P., and Francis, K. (2023). *Data Protection as a Corporate Social Responsibility*, Edward Elgar Publishing.

[5] ENISA (2023). *Multilayer Framework for Good Cybersecurity Practices for AI*. Retrieved April 13, 2024, from https://www.enisa.europa.eu/publications/multilayer-framework-for-good- cybersecurity-practices-for-ai.

[6] Zanfir-Fortuna, G., (2023). *How Data Protection Authorities Are De Facto Regulating AI*, Future of Privacy Forum, 2023, retrieved April 13, 2024, from: https://fpf.org/blog/how-data-protection- authorities-are-de-facto-regulating-generative-ai/.

[7] Ranchordas, S. & Vinci, V. (2024). *Regulatory Sandboxes and Innovation-friendly regulation: Between Collaboration and Capture*. Italian Journal of Public Law, 16(1). Retrieved April 13, 2024, from: https://www.ijpl.eu/wp-content/uploads/2024/03/8.-Ranchordas-and-Vinci.pdf.

[8] Madiega, T. & Louise Van De Pol, A. (2022). *Artificial intelligence act and regulatory sandboxes*. European Parliamentary Research Service. Retrieved from: https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI(2022)733544_EN.pdf.

[9] Federal Ministry for Economic Affairs and Energy (BMWi) (2019). *Making space for innovation: The handbook for regulatory sandboxes*. Retrieved April 13, 2024, from: https://www.bmwk.de/Redaktion/EN/Publikationen/Digitale-Welt/handbook-regulatory- sandboxes.pdf?__blob=publicationFile&v=2.

[10] Datatilsynet (n.d.). *Regulatory privacy sandbox*. Retrieved April 13, 2024, from: https://www.datatilsynet.no/en/regulations-and-tools/sandbox-for-artificial-intelligence/.
[11] Parenti, R. (2020). *Regulatory Sandboxes and Innovation Hubs for FinTech Impact on innovation, financial stability and supervisory convergence*. Policy Department for Economic, Scientific and Quality Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg, DOI: 10.2861/14538.
[12] Undheim, K., Erikson, T., Timmermans, B. (2023). *True uncertainty and ethical AI: regulatory sandboxes as a policy tool for moral imagination*. AI and Ethics, 3, 997–1002, DOI: 10.1007/s43681- 022-00240-x.
[13] Oxford Reference. (n.d.) regulatory capture. In Oxford Reference. Retrieved April 13, 2024, from: https://www.oxfordreference.com/view/10.1093/oi/authority.20110803100411608.
[14] Brown, E., & Piroska, D. (2022). *Governing Fintech and Fintech as Governance: The Regulatory Sandbox, Riskwashing, and Disruptive Social Classification*. New Political Economy, 27(1), 19–32, DOI: 10.1080/13563467.2021.1910645.
[15] Markussen, T. (2023). Evaluation of the Norwegian Data Protection Authority's Regulatory Sandbox for Artificial Intelligence. Datatilsynet. Retrieved April 13, 2024, from: https://www.datatilsynet.no/contentassets/41e268e72f7c48d6b0a177156a815c5b/agenda-kaupang- evaluation-sandbox_english_ao.pdf.
[16] Ranchordas, S., (2021). *Experimental Regulations for AI: Sandboxes for Morals and Mores*. University of Groningen Faculty of Law Research Paper Series, No. 7/2021. DOI: 10.2139/ssrn.3839744.