



Task Force 05

**INCLUSIVE DIGITAL TRANSFORMATION**

## Agile Governance for an Agile Future: Sandboxes for Promoting Responsible Innovation

Giovana Carneiro, Harvard Law School, United States

Thiago Guimarães Moraes, University of Brasilia, Brazil

Carolina Rossini, Datasphere Initiative, Switzerland



**TF05**

## Abstract

This policy brief delves into the escalating concerns surrounding Artificial Intelligence (AI) innovation and proposes a strategic approach to address them. With a foundation in **research-based evidence**, our focus centers on the instrumental **role of sandboxes in fostering responsible AI development and mitigating associated risks**.

**Highlighting the experiences of data protection authorities**, we illustrate how sandboxes uphold **ethical principles throughout the entire technology lifecycle**, emphasizing privacy by design. Beyond their technical functionalities, sandboxes contribute significantly to agile development, ensuring the active participation of diverse stakeholders ranging from innovators to policymakers and communities potentially affected.

This paper emphasizes the importance of collaboration and cooperation in the realm of AI, and demonstrates how employing an interdisciplinary and transnational approach can enable a more holistic understanding of the technology's implications. By facilitating feedback loops with affected communities, sandboxes become catalysts for ethically refining AI technologies and promoting innovation. As we navigate AI's complexities, we evaluate its challenges and opportunities, highlighting the necessity for robust ethical governance frameworks. The proactive incorporation of sandboxes emerges as a crucial measure for ensuring responsible AI development and deployment, aligning seamlessly with SDGs 3, 4, 8, 9, 10, 11, 12, 13, 16, and 17.

**Keywords:** Regulatory sandboxes; Artificial Intelligence; Responsible Innovation; Data Protection

## Diagnosis of the Issue

This policy brief centers on the instrumental role of (regulatory) sandboxes in fostering responsible artificial intelligence development and mitigating associated risks. Highlighting the experiences of data protection authorities (DPA), we illustrate how sandboxes uphold ethical principles throughout the entire technology lifecycle, emphasizing privacy by design. Beyond their technical functionalities, sandboxes contribute significantly to agile development and testing, ensuring the active participation of diverse stakeholders ranging from innovators to policymakers and communities potentially affected (OECD, 2019).

Sandboxes were modeled initially on the enclosed software testing environments that programmers create to test and run code and applications without risking the overall system they work in. This practice was adopted in at least two ways, evolving into operational and regulatory sandboxes.

Operational sandboxes are secure testing environments that enable technology design and testing, including pooling datasets in data-dependent technologies such as AI. Regulators or government agencies can create operational sandboxes to test capabilities on actual datasets, or coalitions of actors can pool resources together, sometimes through technologically enabled decentralized approaches (e.g., data collaboratives, fiduciaries, commons), to explore or encourage their use. Examples of operational sandboxes include a sandbox created by the UK's National Institute for Health Data Services (HDRUK, 2019).

In their turn, regulatory sandboxes are time-limited collaborative endeavors involving regulators, service providers, and other relevant stakeholders to test innovative technology and data practices against regulatory frameworks. Regulatory sandboxes are

not necessarily physical spaces, but rather collaborative processes through which technological innovations and new business models are explored with regulators. Drilling down further into regulatory sandboxes, significant variation exists between them (DATASPHERE, 2022).

Regulatory sandboxes were piloted in the financial sector in the mid-2010s when a desire emerged to evaluate and modernize financial regulations in the advent of fintech. The main goal was regulatory flexibility, temporarily suspending rules to reduce regulatory barriers for entrants and allow a better understanding of new services.

Regulatory sandboxes have also gained traction in the data protection community. One such example is the Norwegian DPA, the “Datatilsynet,” which launched an initiative to promote the development of trustworthy AI. It also seeks to help organizations implement privacy-by-design to enable compliance with the EU GDPR.

DPA's oversee regulatory frameworks based on the risk-based approach supported by accountability principles<sup>1</sup>, in which regulated actors don't need licenses to operate in the market, such as those needed in the financial sector. DPA's sandboxes focus on promoting responsible innovation in alignment with the respective data protection legislation. In essence, they work as "supervised privacy by design" experimentation (CIPL, 2019).

More recently, regulators have also looked into AI sandboxes. The EU AI Act's ambition to promote trustworthy AI innovation and evaluate AI systems within a tiered risk classification structure has incentivized AI regulatory sandboxes.

While there is a common perception that technology regulation impedes innovation (BRADFORD, 2024), sandboxes have the potential to demonstrate how both can be reconciled. They can be a living example that more regulation does not necessarily mean

---

<sup>1</sup> For example, in the Brazilian data protection legislation, Law n. 13.709/2018 (LGPD), the accountability principle is one of its key elements in promoting risk-based regulation.

less innovation. They allow for collaboration, evidence-based, and timely regulation. AI sandboxes are a new mechanism to foster innovation while safeguarding fundamental rights and engraining responsible innovation in the AI development process.

The EU is not the only region that is calling for AI sandboxes. The Brazilian AI Bill<sup>2</sup> also foresees them. However, while the EU AI Act has provided rules on how regulators should develop AI sandboxes and safeguards to fundamental rights, the current version of the Brazilian AI Bill lacks this substance. Other AI regulations may suffer from the same issue—the lack of a shared definition of sandboxes may bring very different rules that could jeopardize their potential to provide collaboration between regulators and innovators.

Furthermore, a lack of expertise in implementing sandboxes may result in failed experiences that discourage stakeholders from implementing this tool. Given this, we share some recommendations for developing robust sandbox programs.

---

<sup>2</sup> Currently, the provisions for regulatory sandboxes in PL 2338/2023 lack a definition on what are regulatory sandboxes, and more detailing on how they should be implemented.

## Recommendations

In this policy brief, we suggest three core but non-exhaustive recommendations for G20 Member States to consider when implementing (regulatory) sandboxes, particularly those that focus on fostering responsible AI development while mitigating associated risks.

**Recommendation 1: AI Sandboxes should not focus on regulatory leeway but rather on fostering a responsible innovation environment.**

**Norms and legislation concerning sandboxes should not solely emphasize regulatory flexibility but should rather prioritize fostering a supervised environment for responsible innovation.** While regulatory leeway is important for enabling experimentation and fostering innovation within sandboxes, it must be balanced with robust oversight mechanisms to ensure that risks are adequately managed and human rights protection is upheld. By focusing on supervised, responsible innovation, regulatory frameworks can promote the development of new technologies and business models while safeguarding against potential harms such as data breaches, financial instability, or ethical concerns. **This approach encourages collaboration between regulators, industry stakeholders, and other relevant parties to establish clear guidelines, risk mitigation strategies, and accountability mechanisms within sandbox environments, ultimately contributing to sustainable and inclusive innovation ecosystems.** For instance, regulatory sandboxes are one of the primary mechanisms by which the EU AI Act will encourage and promote trustworthy AI innovation.

A closer look at the versions of the EU AI Act is helpful to illustrate this normative difference. The first draft provisions on sandboxes had a stronger focus on innovation. In contrast, the version approved in Parliament included various safeguards. For instance, it highlights the supervisory and corrective powers of the agencies involved, including an explicit mention of the existing liability regime for third-party damages as a result of experimentation (Article 57(12)). In short, the final text of the EU AI Act presents a *responsible innovation* approach for regulatory sandboxes.

**Recommendation 2: AI sandboxes should take guidance from privacy sandboxes, which have applied “by design” approaches to ethics and human rights**

Several privacy sandboxes, such as those hosted by DPAs, could be considered benchmarks for the development of ethical AI sandboxes. For example, Singapore’s PDPC and the UK’s ICO have been conducting privacy sandboxes since 2017 and 2019, respectively. Although these experimentations did not focus on AI, some AI systems have been selected as use cases. For the ICO, sandboxes have provided valuable outputs, such as its guidance on AI and data protection (ICO, 2023).

Alternative approaches have been observed in Colombia’s SIC, Norway’s Datatilsynet, and France’s CNIL, which decided to focus their sandboxes specifically on AI-driven technologies. All of them highlighted the importance of developing reliable AI systems that observe the implementation of ethical values such as those fostered by privacy by design. In the case of SIC (2021), privacy by design and by default was explicitly mentioned as the project's main goal.

**Recommendation 3: G20 Member States should cooperate when designing, planning, and implementing AI sandboxes**

Regulatory sandboxes take a lot of time and effort to design, plan, and implement. While numerous countries have documented their experiences or enacted laws enabling the establishment of sandboxes, regulators often require additional expertise and practical experience to leverage these frameworks effectively. **The G20 can be a forum for cooperation between Members when developing such a promising mechanism.** The existing network of regulators can be a space to promote this collaboration with competent authorities exchanging best practices and sharing insights and lessons learned from their own experiences. When implementing a sandbox for the first time, the sandbox itself can be an experiment.

Alternatively, if the G20 does not wish to create this space anew, **ongoing efforts such as the Datasphere Initiative's Global Sandboxes Forum (DATASPHERE, 2024) or other efforts by their Members may be leveraged and/or expanded.** For instance, the Brazilian DPA has engaged in an extensive benchmark effort and considered the experience of other data protection agencies around the globe when developing their own sandboxes. This experience can be pivotal to other G20 countries. Learning from ongoing efforts can also leverage processes' best practices and improve accountability and transparency.

Moreover, **forums organized by think tanks with various public and private stakeholders can promote space for meaningful collaboration.** Initiatives like the African Forum on Sandboxes for Data and the Global Sandboxes Forum are compelling examples of how regulators can come together to cultivate the necessary expertise. In that sense, the Datasphere Initiative is developing a sandbox roadmap and evaluation



framework to guide policymakers. By participating in such forums, regulators and other stakeholders can share insights, best practices, and lessons learned from sandbox implementations across various jurisdictions. These collaborative efforts facilitate the exchange of knowledge and foster the development of expertise needed to navigate the complexities of regulating new technologies and business models within sandbox environments.

### Scenario of Outcomes

Based on the recommendations above, we foresee various positive outcomes from the adoption of sandboxes while pointing out potential trade-offs and challenges.

#### Outcome 1: Risk Mitigation

Regulators can use sandboxes to test new regulations or regulatory frameworks in a controlled environment before implementing them on a larger scale. This helps identify potential risks and challenges before they impact the broader market.

On the other hand, there might be a **trade-off** as sandboxes influence the *timing* for the private sector to roll out certain innovations. Provided that there are enough incentives for a private player to participate in the sandbox, it also has to be mindful that it will *test* the product within a controlled environment before making it available to the general public.

#### Outcome 2: Regulators are better equipped to understand innovation with increased *expertise*.

Sandboxes allow regulators to understand better emerging innovative technologies and business models in their industries. In other words, they give regulators firsthand contact

with the latest technological developments and solutions. By engaging with companies operating within the sandbox, regulators can stay abreast of developments and tailor regulations to support innovation while ensuring consumer protection and market integrity.

Moreover, considering the profound evolution of our societies and economies amid the fourth industrial revolution, each nation and its governing bodies must collaborate to facilitate this digital shift and harness its advantages while mitigating potential risks. Sandboxes are core tools supporting this exercise.

Overall, sandboxes offer regulators a valuable tool for keeping up with technological advancements, fostering innovation, and ensuring that regulatory frameworks remain effective in a rapidly evolving landscape.

### **Outcome 3: Meaningful participation and collaboration between different stakeholders**

One of the main goals of regulatory sandboxes is to promote collaboration between several stakeholders to test and learn how their innovations can be compatible with regulatory systems. Adopting these recommendations can result in meaningful participation and collaboration among all parties.

Sandboxes provide a platform for collaboration between regulators, industry participants, and other stakeholders. By working together within the sandbox environment, regulators can establish closer relationships with industry players, share knowledge, and develop more effective regulatory approaches.

In a sandbox, private entities and public authorities are aligned and aware of its rules, goals, and possibilities. One important and understandable **concern** from private entities is the potential to be sanctioned for violations of rules within the sandbox. Whether or not

it will involve the suspension of norms—including those related to sanctioning—will depend on (i) the level of authority an agency has to suspend sanctions and (ii) the regulator’s discretion regarding the trade-offs involving such a waiver. For instance, it is unlikely that any jurisdiction will allow the authority to suspend the application of criminal laws. Similarly, a regulator might choose not to suspend a certain rule that protects a right so fundamental that it should not be violated in any circumstance—not even in a testing environment.

Aware of such a challenge, Article 57(7) of the EU AI Act states that competent authorities aiming to create a sandbox “using the AI regulatory sandbox with guidance on regulatory expectations” and, following terms and conditions for the providers’ participation, “no administrative fines shall be imposed by the authorities” for infringements of that same regulation (Article 57(12)). This provision safeguards the level of trust among the stakeholders involved and clarifies that the sandbox is not a trap for auditing and fining private entities. On the contrary, it is a place for mutual agreement and clear expectations.

#### **Outcome 4: Facilitating Regulatory Development, Understanding, and Compliance**

Sandboxes enable regulators to work closely with companies and impacted communities to ensure compliance with existing regulations and assist in navigating regulatory requirements in innovative areas. This collaborative approach can help companies better understand their regulatory obligations and minimize compliance-related issues.

On the other hand, an important **risk** is a lack of transparency. We do not see it as a **trade-off** but rather as an obstacle to consider when planning and executing sandboxes. Even when the G20 Member States achieve full engagement within and between

sandboxes (at national and international scales), the experimentation must be as transparent as possible. This is necessary to build trust with individuals and society since they should be the main beneficiaries of these experiments. If a sandbox is not developed with the aim of public interest and the greater good, its goals should be revised.

One way to achieve transparency is by publishing reports that share the lessons learned from the sandbox experimentation. Privacy regulators such as the ICO (2024) and Datatilsynet (2024) have consistently published their reports of past experiences. One point of attention is that these reports have to be careful not to violate IP rights and other trade secrets since this would hinder the private sector's interest in participating in further experimentations. However, the experiences mentioned here prove that this balance can be achieved. Furthermore, knowledge sharing can go even further than reports. Datatilsynet has developed other approaches, such as podcasts and webinars, to share lessons learned on its sandboxes with a broader public.

## **Conclusion**

Sandboxes should benefit all stakeholders involved. They prepare regulators to provide guidelines for legal compliance and align innovators with ethical principles, such as privacy and data protection. The main outcome of a properly implemented sandbox is a society where trustworthy innovations can thrive.

## References

Bradford, Anu, The False Choice Between Digital Regulation and Innovation (Bradford, 2024). Northwestern University Law Review, Vol. 118, Issue 2, October 6, 2024,

Available at [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4753107](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4753107)

Center for Information and Policy Leadership - CIPL, ‘Regulatory Sandboxes in Data Protection: Constructive Engagement and Innovative Regulation in Practice’ (CIPL, 2019)

<[https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl\\_white\\_paper\\_on\\_regulatory\\_sandboxes\\_in\\_data\\_protection\\_-\\_constructive\\_engagement\\_and\\_innovative\\_regulation\\_in\\_practice\\_\\_8\\_march\\_2019\\_.pdf](https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_white_paper_on_regulatory_sandboxes_in_data_protection_-_constructive_engagement_and_innovative_regulation_in_practice__8_march_2019_.pdf)> accessed 16 March 2024

Datatilsynet, ‘Sandbox - reports’ (Datatilsynet, 2024)

<<https://www.datatilsynet.no/en/regulations-and-tools/reports-on-specific-subjects/>> accessed 16 March 2024

Commission Nationale de l’Informatique et des Libertés – CNIL, ‘“Sandbox”: CNIL launches call for projects on artificial intelligence in public services’ (CNIL, 2022)

<<https://www.cnil.fr/en/sandbox-cnil-launches-call-projects-artificial-intelligence-public-services>> accessed 16 March 2024.

Datatilsynet, ‘Sandbox for Responsible Artificial Intelligence’ (Datatilsynet, 2021)

<<https://www.datatilsynet.no/en/regulations-and-tools/sandbox-for-artificial-intelligence/>> accessed 16 March 2024.

European Union - EU. EU AI Act (EU, 2024)

<<https://www.euaiact.com/article/3#:~:text='AI%20regulatory%20sandbox'%20means%20a,specific%20plan%20for%20a%20limited>> accessed 27 March 2024.

Health Data Research UK - HDRUK. Call for applications to test new £37.5M UK health data research services. (HDRUK, 2019) <<https://www.hdruk.ac.uk/about-us/what-we-do/health-data-research-explained/>> accessed 27 March 2024.

ICO, 'Regulatory Sandbox - Previous Participants' (ICO, 2024) <<https://ico.org.uk/for-organisations/regulatory-sandbox/previous-participants/>> accessed 16 March 2024

Information Commissioner's Office – ICO, 'Guidance on AI and data protection' (ICO, 2023) <<https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/>> accessed 16 March 2024.

Information Commissioner's Office – ICO, 'Regulatory Sandbox' (ICO, 2019) <<https://ico.org.uk/sandbox>> accessed 16 March 2024

Organization for Economic Cooperation and Development - OECD, The role of sandboxes in promoting flexibility and innovation in the digital age (2019).

<[https://goingdigital.oecd.org/data/notes/No2\\_ToolkitNote\\_Sandboxes.pdf](https://goingdigital.oecd.org/data/notes/No2_ToolkitNote_Sandboxes.pdf)> accessed 27 March 2024.

Personal Data Protection Commissioner – PDPC, 'A Trusted Ecosystem for Data Innovation' (PDPC 2017) <[https://www.pdpc.gov.sg/-/media/Files/PDPC/New\\_DPO\\_Connect/aug\\_2017/pdf/ATrustedEcosystemForDataInnovation.pdf](https://www.pdpc.gov.sg/-/media/Files/PDPC/New_DPO_Connect/aug_2017/pdf/ATrustedEcosystemForDataInnovation.pdf)> accessed 16 March 2024.

Superintendencia De Industria Y Comercio – SIC, 'Sandbox on Privacy by Design and by Default in Artificial Intelligence' (SIC, 2021)

<<https://www.sic.gov.co/sites/default/files/files/2021/150421%20Sandbox%20on%20privacy%20by%20design%20and%20by%20default%20in%20AI%20projects.pdf>> accessed 16 March 2024.

The Datasphere Initiative, 'Global Sandboxes Forum' (Datasphere, 2024)

<<https://www.thedatasphere.org/programs/lab-for-innovation/sandboxes/>> accessed 27 March 2024

The Datasphere Initiative, 'Sandboxes for Data: Creating Spaces for Agile Solutions across Borders' (Datasphere, 2022) <<https://www.thedatasphere.org/datasphere-publish/sandboxes-for-data/>> accessed 27 March 2024



# Let's **rethink** the world

