REGULATORY SANDBOX

AI for the amicable resolution of administrative conflicts at the Attorney General's Office of Brazil

2025

Labori
Innovation Laboratory of the AGU

AGU
ATTORNEY GENERAL'S OFFICE

LABORATORY OF INNOVATION
ATTORNEY GENERAL'S OFFICE OF BRAZIL

Bruno Portela
Team Lead

Leila de Morais
Deputy Team Lead

Bruno Andrade Costa

Daniel Pereira de Franco

Eliza Victoria Silva Lemos

Julia Correa da Silva Araújo

Michele Cristina Silva Melo

Robson Ramos da Silva

Adriana Cristina de Paula

Nadya Rodovalho Souza Galera

Mauro César Santiago Chaves

Key Contributors

Alexandra da Silva Amaral

Carlos Henrique Benedito Nitão Loureiro

Luiz Fabrício Thaumaturgo Vergueiro

Adalberto Maciel

Carlos Octaviano de Medeiros Mangueira

Leopoldo Gomes Muraro

Jezihel Pena Lima

Mariana Cruz Montenegro

Henrique Tróccoli Júnior

Adriana Macedo Marques

Tatiana Meinhart Hahn

Flávia Correa Azeredo de Freitas

Teresa Villac Pinheiro

Diego Pereira

Fernanda Rodrigues de Morais

Frederico Rios Paula

Priscila Gonçalves de Oliveira

Ricardo Cavalcante Barroso

Luis Eduardo Sales Cordeiro

James Castelo Branco Costa Filho

Caio Marcio Melo Barbosa

Francisco Alexandre Colares Melo Carlo

LABORI NOTE

The Innovation Laboratory of the Attorney General's Office of Brazil (LABORI/AGU) is pleased to present this study on the design of a regulatory sandbox for the adoption of artificial intelligence (AI) in the mediation of administrative conflicts.

The study explores, in an experimental and forward-looking manner, how a controlled regulatory environment could support the consensual resolution of administrative disputes, particularly those arising from the denial of social security benefits. The proposal seeks to reduce unnecessary judicialization and promote greater efficiency in the performance of the State, while reinforcing legal certainty and public trust.

More than a prototype of technological innovation, the sandbox is analyzed here as an instrument of institutional learning and experimental regulatory governance, capable of generating evidence that can inform the improvement of public policies and guide future decisions on the use of AI in sensitive legal functions.

This study reaffirms the role of the AGU as an institution committed to responsible innovation, administrative modernization, and the development of experimental regulatory solutions, contributing to a more efficient, transparent, and citizen-oriented public administration.

INTERNATIONAL PERSPECTIVE

The adoption of artificial intelligence (AI) across the public sector has often been approached with caution. Early misuses, such as biased decision-making tools documented almost a decade ago, fueled the perception that deploying AI in sensitive domains was too risky. Yet this perception risks obscuring the enormous potential of AI to improve efficiency, strengthen public services, and enhance trust in institutions, provided the technology is tested in real conditions with robust safeguards. Regulatory sandboxes make this possible by offering controlled environments where innovation can be responsibly introduced, evaluated, and refined.

Brazil's Attorney General's Office (AGU), through its Innovation Laboratory (LABORI), is leading one of the most advanced and innovative sandbox initiatives in the region. This effort focuses on the consensual resolution of administrative disputes, particularly those arising from denied social security benefits that too often drive unnecessary judicialization. By intervening earlier in the process, the initiative aims to reduce the burden on the judiciary while reinforcing legal certainty and strengthening public confidence. Equally important, it allows authorities to determine where AI can serve as a valuable tool and where human oversight must remain indispensable.

The implications extend beyond Brazil. This initiative provides a model for addressing pressing questions of governance: how to ensure transparency, safeguard privacy, and define standards for human intervention in the use of AI, including more advanced systems such as agentic AI. For Brazil, the sandbox is more than a mechanism for experimentation; it is a laboratory for institutional learning, helping to shape governance arrangements adapted to national needs.

It has been a privilege to contribute to this initiative as both author and technical expert, working alongside AGU with the support of GIZ and ECLAC. Developed in record time through the RESMA methodology, the project shows that rigorous sandbox design and practical implementation can go hand in hand, offering a blueprint for future efforts in Brazil and beyond.

Armando Guio

SUMMARY

REGULATORY SANDBOX

AI for the amicable resolution of administrative conflicts at the Attorney General's Office of Brazil

INSTITUTION: ATTORNEY GENERAL'S OFFICE OF BRAZIL (AGU)

SUBJECT: ESTABLISHES THE INTERNAL FRAMEWORK FOR THE IMPLEMENTATION AND GOVERNANCE OF A REGULATORY SANDBOX AIMED AT TESTING ARTIFICIAL INTELLIGENCE (AI) TECHNOLOGIES IN THE AMICABLE RESOLUTION OF ADMINISTRATIVE CONFLICTS.

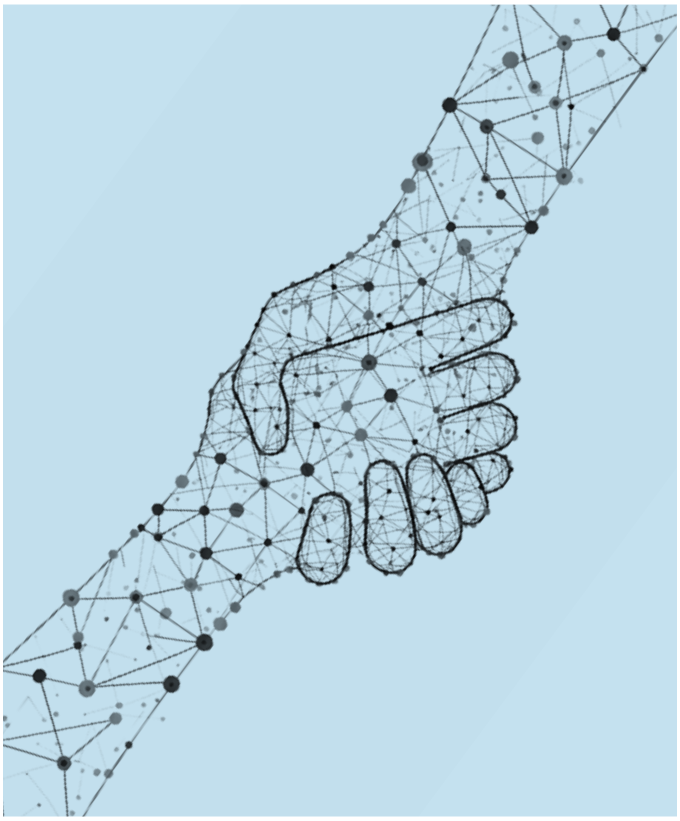
AUTHORS: ELIZA VICTÓRIA LEMOS, FEDERAL ATTORNEY, ATTORNEY GENERAL'S OFFICE OF BRAZIL (AGU)
ARMANDO GUIO - GIZ AND ECLAC CONSULTANT

Labori
Innovation Laboratory of the AGU

AGU
ATTORNEY GENERAL'S OFFICE

CONTEXT

The Attorney General's Office of Brazil (AGU) is currently developing an artificial intelligence (AI) system designed to support the amicable resolution of administrative disputes—particularly those arising from the denial of social security benefits by the National Institute of Social Security (INSS). In view of the complex legal, ethical, and operational considerations involved in deploying AI in public decision-making, this proposal outlines the creation of a regulatory sandbox: a structured, temporary, and controlled experimental environment that facilitates testing, evaluation, and institutional learning.



This document provides a detailed roadmap for the sandbox's design and operationalization. It articulates both the rationale and methodological foundations underpinning the initiative, identifies anticipated challenges, and delineates preparatory steps to ensure institutional readiness and regulatory alignment. It also aims to enhance transparency and stakeholder engagement by framing the sandbox as a public governance innovation with significant implications for the future of AI in Brazil's administrative state.

Importantly, this proposal marks the first application of the Regulatory Sandbox Maturity Assessment (RESMA) methodology within a Latin American public institution[1]. RESMA, an emerging international benchmark, offers a rigorous framework to assess the legal, institutional, and technical readiness of public bodies to implement experimental regulatory mechanisms. The findings of this initiative will not only shape the evolution of AI governance in Brazil, but also contribute to the global refinement of RESMA itself as a policy tool.

[1] Guio, Armando. Regulatory Sandboxes in Developing Economies: An Innovative Governance Approach. Santiago, Chile: Economic Commission for Latin America and the Caribbean (ECLAC), July 19, 2024. https://repositorio.cepal.org/entities/publication/4e066e90-ea1a-454a-b266-d2e6da90abba.

ALIGNMENT WITH PUBLIC POLICIES AND INSTITUTIONAL STRATEGIES

The AGU's AI initiative is deeply aligned with national and institutional strategies that prioritize digital transformation, legal innovation, and administrative modernization. It reinforces the core objectives of the AGU's Institutional Development Plan and its Strategic Innovation Agenda, which emphasize the promotion of legal certainty, efficiency in public service delivery, and the responsible adoption of frontier technologies.

At the national level, the project advances key priorities set forth in the Brazilian National Artificial Intelligence Strategy (Estratégia Brasileira de Inteligência Artificial), which encourages safe experimentation with AI under guiding principles of ethics, transparency, and accountability. Through its experimental governance model, the sandbox provides a risk-controlled environment for developing best practices around AI-assisted legal functions.

Additionally, the initiative supports the goals of the Brazilian Digital Government Strategy (Estratégia de Governo Digital) by enabling digital innovation in legal processes, increasing institutional agility, and improving access to services. By integrating AI tools into legal workflows, the AGU moves decisively toward a more efficient and data-informed public administration.

The project also demonstrates strict adherence to the Brazilian General Data Protection Law (Lei Geral de Proteção de Dados - LGPD). It incorporates a robust data governance framework grounded in principles such as privacy by design, purpose limitation, and continuous auditability.

Finally, the sandbox complements the National Policy for the Improvement of Public Administration and the broader Regulatory Improvement Agenda, both of which call for flexible regulatory mechanisms to accommodate innovation. In this sense, the project serves not only as a technology initiative but also as a prototype for legal institutional transformation that enhances democratic legitimacy and public trust.

IMPACT ON BRAZIL'S AI GOVERNANCE

This regulatory sandbox is conceived as a foundational pilot to support the responsible integration of artificial intelligence in the public sector. Its core objective is to illustrate how experimental governance tools can facilitate the safe and legitimate adoption of AI in areas often considered too sensitive or high-risk—such as social protection, legal services, and access to justice.

Rather than defaulting to prohibitions on AI deployment in such domains, the sandbox is designed to identify, assess, and mitigate associated risks. It will test strategies to reconcile technological innovation with constitutional and legal safeguards, with particular emphasis on challenges such as automation bias, explainability, and the design of appropriate human oversight—each of which becomes more salient as AI systems increase in complexity and autonomy.

One particularly relevant development is the rise of agentic AI—systems capable of pursuing long-term objectives through autonomous actions, even when specific behaviors have not been explicitly programmed in advance. The degree of agenticness in a system may be understood as its capacity to flexibly and effectively achieve complex goals in dynamic environments with minimal direct supervision[2].

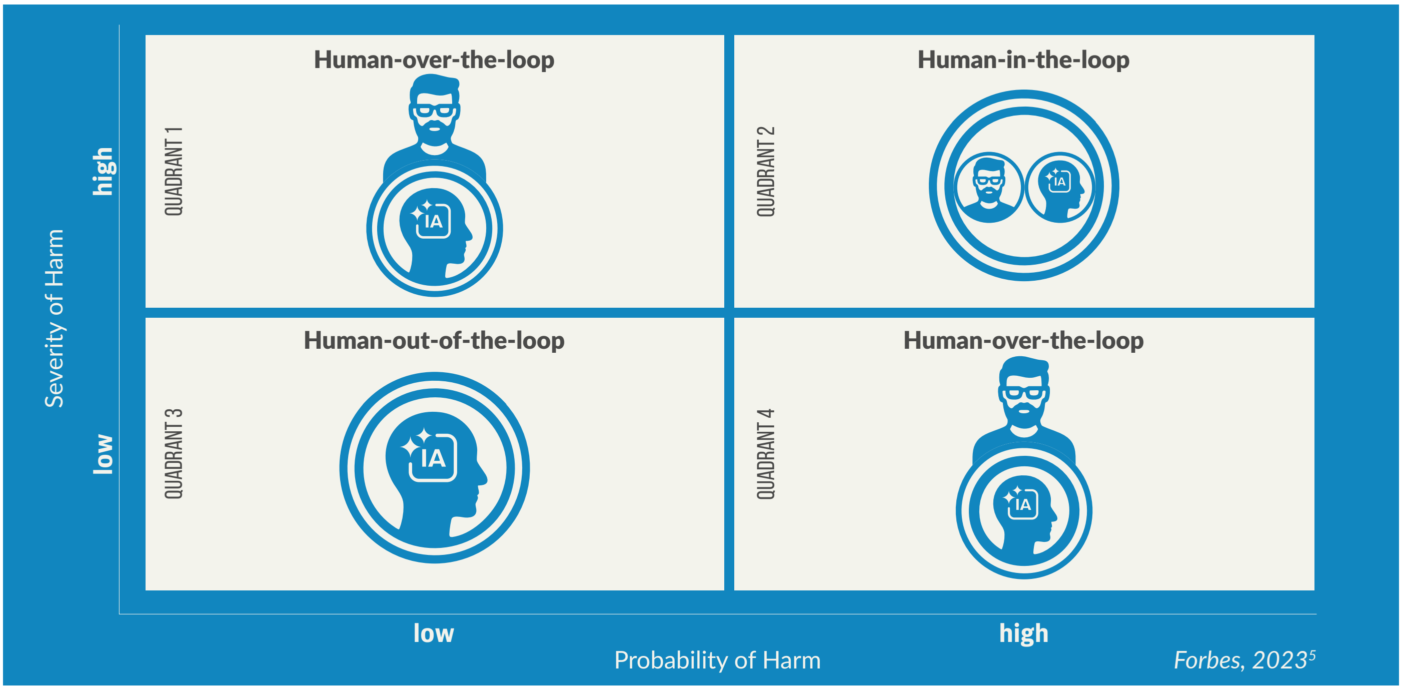
Understanding the implications of agentic AI in public administration is essential, especially in defining the role of public officials as potential users or supervisors of such systems. The user of an agentic AI system is typically the individual or institution that activates it, sets its operational objectives, and exercises a degree of oversight over its outputs and interactions. During operation, these systems may engage with third parties—including other humans or digital services—depending on the tasks assigned.

[2] Yonadav Shavit et al. Practices for Governing Agentic AI Systems. San Francisco: OpenAI, 2024. *https://cdn.openai.com/papers/practices-for-governing-agentic-ai-systems.pdf*.

In their capacity as users or implementers, public officials must be able to evaluate whether a given agentic AI system is appropriate for the intended function. This includes determining whether the system can reliably perform across expected operational conditions. Where full reliability is not essential due to the low-risk nature of a task, interface design and user guidance become especially important to ensure transparency and alignment of expectations. These considerations underscore the urgent need to establish robust methodologies for evaluating agentic AI systems, including the capacity to anticipate and respond to potential failure modes[3]. In this regard, the sandbox presents a valuable opportunity to generate empirical evidence that can inform broader evaluations of such systems—particularly as they are introduced into critical areas of Brazil's public administration.

A central focus of the sandbox will be to delineate the role of human officials in AI-supported decision-making. This includes establishing clear parameters for discretion, supervisory responsibilities, and mechanisms of accountability.

Through iterative testing, the AGU will explore various models of human-in-the-loop oversight and evaluate their implications for administrative legitimacy and public trust. A particularly noteworthy reference point is Singapore, whose approach has attracted considerable international attention. Its model calibrates the degree of human oversight based on the severity of potential harm and the likelihood of its occurrence. Accordingly, different levels of human control are envisioned for AI-automated activities, with the aim of ensuring the most effective and proportionate form of supervision. In low-risk scenarios—where both the probability and impact of harm are minimal—human review may be unnecessary. Conversely, in high-stakes contexts, active and robust human intervention becomes essential[4].



Forbes, 2023[5]

[3] Ibid.

[4] Personal Data Protection Commission (PDPC). Model AI Governance Framework (Second Edition). Singapore: PDPC, January 21, 2020. *https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf.*

In the Brazilian context, the objective is to adapt and refine these concepts within institutions such as AGU, with a focus on clarifying how harm should be defined—its nature, degrees of severity, and potential consequences for public decision-making. This regulatory sandbox will offer a structured environment in which to develop and test these evaluative criteria, thereby informing the safe and principled deployment of autonomous systems across domains such as public service delivery, citizen engagement, and even judicial processes. The insights generated through this regulatory experiment will play a critical role in shaping Brazil's broader AI governance framework, particularly with respect to standards for human oversight.

Finally, the sandbox is expected to provide concrete insights relevant to legislative initiatives such as Bill No. 2338/2023, currently under debate in the Brazilian Congress. This bill mandates strong human oversight for high-risk AI systems, including requirements that:

- Users understand the system's capabilities and limitations and can maintain effective operational control;
- Users are aware of the risks of automation bias and undue reliance on AI outputs;
- Users are equipped to interpret AI-generated outcomes in light of the system's architecture and applicable interpretive tools.

By grounding AI regulation in practical, real-world experimentation, this initiative affirms a foundational principle of democratic innovation: that regulatory frameworks should evolve in tandem with technological development—not in isolation from it.

5. Gow, Glenn. 2023. "A Simple AI Governance Framework in the Age of ChatGPT." Forbes, August 6, 2023. *https://www.forbes.com/sites/glenngow/2023/08/06/a-simple-ai-governance-framework-in-the-age-of-chatgpt/.*

OBJECTIVES OF THE SANDBOX

The primary goal of the sandbox is to test the AI system's functionality within a real yet controlled environment, allowing the AGU to:

- Analyze regulatory, legal, and ethical implications of AI-assisted legal tasks;
- Evaluate the operational feasibility of optimizing internal workflows;
- Enhance institutional accountability and legal predictability;
- Generate empirical evidence to inform future regulation and digital transformation strategies.

INITIAL USE CASE: INSS BENEFIT DENIALS

The first use case involves administrative disputes stemming from INSS benefit denials. The AI system will:

- Detect patterns among denied claims and highlight potentially eligible cases;
- Generate recommendations and legal risk indicators for internal review;
- Support the negotiation and drafting of extrajudicial agreements;
- Automate the generation of internal legal documents;
- Serve as a prototype for future applications in legal risk prevention and resolution.

PROPOSED GOVERNANCE STRUCTURE

The sandbox will be coordinated by an **Oversight Committee**, comprising:

THE INNOVATION LABORATORY (LABORI/AGU);

THE SECRETARIAT FOR GOVERNANCE AND STRATEGIC MANAGEMENT (SGE/AGU).

This committee will supervise hypothesis validation, authorize procedural flexibilities, and make final decisions regarding the scaling, revision, or termination of tested solutions. Other AGU units or external experts — legal, academic, or technical—may be invited to contribute as appropriate.

DURATION AND METHODOLOGY

The sandbox will unfold over a 12-month period, structured as follows:

STAGE 1: INTERNAL PLANNING (2–3 MONTHS):

Definition of objectives, hypotheses, risk protocols, and legal mappings.

STAGE 2: TRAINING AND PREPARATION (1–2 MONTHS):

Capacity-building for participating teams; sandbox protocols implemented.

STAGE 3: EXPERIMENTATION AND TESTING (6–7 MONTHS):

Execution of use case testing with continuous monitoring and feedback loops.

STAGE 4: EVALUATION AND LEARNING (2–3 MONTHS):

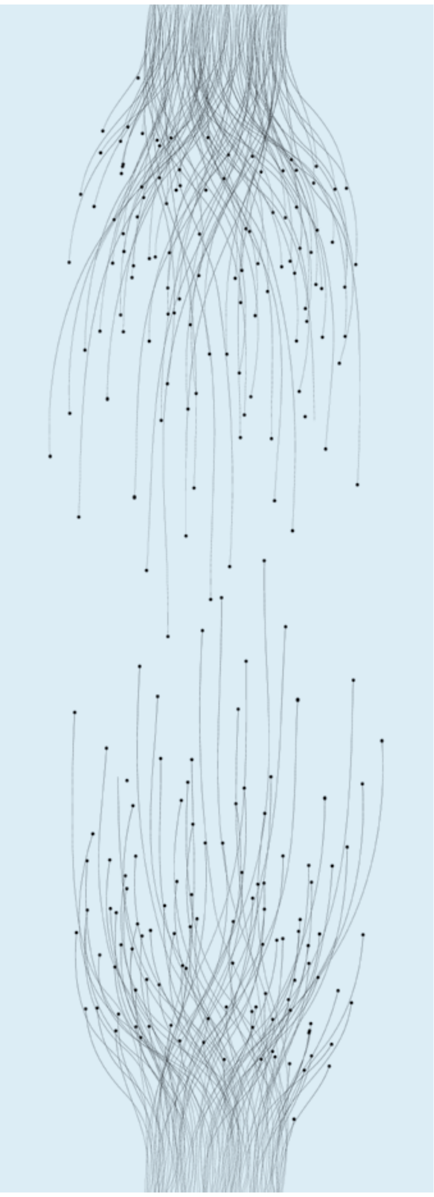
Comprehensive performance evaluation and formulation of policy recommendations.

REGULATORY HYPOTHESES

Potential hypotheses to be tested include:

- Legitimacy of AI-generated legal reasoning;
- Institutional acceptance of AI-assisted legal drafting;
- Lawfulness of data use for model training and validation;
- Role of AI in fostering amicable, out-of-court settlements;
- Explainability and auditability of AI system outputs.

REGULATORY FLEXIBILITIES



Temporary regulatory exceptions may be applied during the sandbox period, including:

- Use of AI for drafting legal documents, suspending exclusive manual drafting requirements;
- Adjusted internal deadlines to accommodate iterative refinement
- Use of experimental authentication tools in compliance with AGU's security policies;
- Pseudonymized or anonymized data use under POSIN-AGU protocols;
- Controlled adjustments to system access and traceability under strict oversight;
- Provisional regulatory regime applicable only within the sandbox;
- Tailored documentation and audit protocols ensuring accountability and reversibility.

All flexibilities will observe fundamental principles of legality, due process, personal data protection, institutional transparency, and revocability in case of identified risk.

MONITORING AND EVALUATION

Robust monitoring will ensure continuous compliance with legal, ethical, and data protection standards. Real-time evaluation mechanisms will track system outputs, detect anomalies, and support timely corrective measures. Evaluation criteria will emphasize institutional integrity, technical performance, and stakeholder trust.

FINAL DELIVERABLES

The Oversight Committee will issue a Final Evaluation Report that includes:

- A technical and legal performance assessment;
- A synthesis of regulatory insights and legal learnings;
- Recommendations for institutional adoption, revision, or discontinuation;
- Drafts of normative instruments as needed for institutionalization.

FINAL REMARKS

This initiative was made possible through the generous support of GIZ and ECLAC and marks a pioneering application of the RESMA methodology. Among the project's key contributions are:

- The ability to predefine the conditions for a structured regulatory experiment;
- The precision with which regulatory priorities were identified and tested;
- The substantial impact on Brazil's AI governance landscape and the value of the documented outcomes.

As one of the first globally documented implementations of RESMA, the lessons from this sandbox offer significant insight into the future of AI regulatory experimentation. The results provide compelling evidence of improved efficiency, shorter processing times, and more effective resource allocation. This experience strongly supports the continued development of experimental governance mechanisms in the public sector.

Labori

Innovation Laboratory of the AGU

AGU

ATTORNEY GENERAL'S OFFICE