

SANDBOX, REGULATÓRIO

Adoção de IA na mediação de
conflitos administrativos na
Advocacia-Geral da União

2025

LABORATÓRIO DE INOVAÇÃO DA ADVOCACIA-GERAL DA UNIÃO

Bruno Portela

Coordenador

Leila de Moraes

Coordenadora Substituta

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

Colaboradores

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 Moraes

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

NOTA DO LABORI

O Laboratório de Inovação da Advocacia-Geral da União (LABORI/AGU) tem a satisfação de apresentar este estudo sobre a concepção de um sandbox regulatório voltado à adoção de inteligência artificial (IA) na mediação de conflitos administrativos.

O estudo busca explorar, em caráter experimental e prospectivo, de que forma um ambiente regulatório controlado poderia apoiar a resolução consensual de controvérsias administrativas, em especial nos casos de indeferimento de benefícios previdenciários. A proposta visa reduzir a judicialização desnecessária e promover maior eficiência na atuação estatal, ao mesmo tempo em que reforça a segurança jurídica e a confiança social.

Mais do que um protótipo de inovação tecnológica, o sandbox é analisado aqui como instrumento de aprendizagem institucional e governança regulatória experimental, capaz de gerar evidências úteis para o aperfeiçoamento das políticas públicas e para orientar futuras decisões sobre o uso de IA em funções jurídicas sensíveis.

Este estudo reforça o papel da AGU como instituição comprometida com a inovação responsável, a modernização administrativa e a construção de soluções regulatórias experimentais, contribuindo para um ambiente público mais eficiente, transparente e próximo da sociedade.



PERSPECTIVA INTERNACIONAL

A adoção da inteligência artificial (IA) no setor público tem sido frequentemente acompanhada de cautela. Os primeiros usos indevidos, como ferramentas de tomada de decisão enviesadas, documentadas há quase uma década, alimentaram a percepção de que aplicar IA em domínios sensíveis seria demasiado arriscado. No entanto, essa visão corre o risco de obscurecer o enorme potencial da IA para melhorar a eficiência, fortalecer os serviços públicos e aumentar a confiança nas instituições, desde que a tecnologia seja testada em condições reais e com salvaguardas robustas. Os sandboxes regulatórios tornam isso possível ao oferecer ambientes controlados nos quais a inovação pode ser introduzida, avaliada e aprimorada de forma responsável.

A Advocacia-Geral da União (AGU), por meio de seu Laboratório de Inovação (LABORI), lidera uma das iniciativas de sandbox mais avançadas e inovadoras da região. O esforço concentra-se na resolução consensual de conflitos administrativos, especialmente aqueles decorrentes da negativa de benefícios previdenciários, que com frequência alimentam a judicialização desnecessária. Ao intervir mais cedo no processo, a iniciativa busca reduzir a sobrecarga do Judiciário, reforçar a segurança jurídica e fortalecer a confiança pública. Igualmente importante, permite identificar em quais áreas a IA pode servir como ferramenta valiosa e onde a supervisão humana deve permanecer indispensável.

As implicações vão além do Brasil. A iniciativa oferece um modelo para enfrentar questões urgentes de governança: como assegurar transparência, proteger a privacidade e definir padrões para a intervenção humana no uso da IA, incluindo sistemas mais avançados, como a IA agente. Para o Brasil, o sandbox é mais que um mecanismo de experimentação; é um laboratório de aprendizado institucional, capaz de moldar arranjos de governança adaptados às necessidades nacionais.

Foi um privilégio contribuir para esta iniciativa como autor e especialista técnico, ao lado da equipe da AGU, com o apoio da GIZ e da CEPAL. Desenvolvido em tempo recorde por meio da metodologia RESMA, o projeto demonstra que é possível unir rigor no desenho de sandboxes e ênfase na implementação prática, oferecendo um modelo para esforços futuros no Brasil e além.

Armando Guio



SUMÁRIO

NOTA DO LABORI	3
PERSPECTIVA INTERNACIONAL	4
CONCEITO	7
ALINHAMENTO COM POLÍTICAS PÚBLICAS E ESTRATÉGIAS INSTITUCIONAIS	8
O IMPACTO NA GOVERNANÇA DE IA NO BRASIL	9
OBJETIVOS DO SANDBOX	12
IMPLEMENTAÇÃO PILOTO: INDEFERIMENTOS DE BENEFÍCIOS DO INSS	13
ESTRUTURA DE GOVERNANÇA	14
DURAÇÃO E METODOLOGIA	15
HIPÓTESES REGULATÓRIAS	16
FLEXIBILIDADES REGULATÓRIAS	17
MONITORAMENTO E AVALIAÇÃO	18
ENTREGAS FINAIS	19
CONSIDERAÇÕES FINAIS	20

SANDBOX REGULATÓRIO

Adoção de IA na mediação de
conflitos administrativos na
Advocacia-Geral da União

INSTITUIÇÃO: ADVOCACIA-GERAL DA UNIÃO (AGU)

ASSUNTO: ESTABELECE A ESTRUTURA INTERNA PARA A IMPLEMENTAÇÃO E GOVERNANÇA DE UM SANDBOX REGULATÓRIO VOLTADO À EXPERIMENTAÇÃO DE TECNOLOGIAS DE INTELIGÊNCIA ARTIFICIAL (IA) NA RESOLUÇÃO AMIGÁVEL DE CONFLITOS ADMINISTRATIVOS.

AUTORES: ELIZA VICTÓRIA LEMOS (ADVOGADA DA UNIÃO/AGU)
ARMANDO GUIO - CONSULTOR GIZ E CEPAL

CONCEITO

A Advocacia-Geral da União (AGU) está em processo de desenvolvimento de um sistema de inteligência artificial (IA) voltado ao apoio na resolução consensual de controvérsias administrativas — notadamente aquelas oriundas da negativa de benefícios previdenciários pelo Instituto Nacional do Seguro Social (INSS).



Considerando a complexidade jurídica, ética e operacional inerente à aplicação de sistemas de IA na tomada de decisão pública, propõe-se a criação de um sandbox regulatório: ambiente experimental, estruturado, temporário e controlado, destinado à realização de testes, avaliações e à promoção do aprendizado institucional.

O presente documento apresenta um roteiro detalhado para a concepção e a operacionalização do sandbox regulatório, contemplando os fundamentos e a metodologia da iniciativa, a identificação dos principais desafios e a definição das etapas preparatórias necessárias à garantia da prontidão institucional e da conformidade regulatória. Busca-se, ainda, fomentar a transparência e o engajamento dos diversos atores envolvidos, posicionando o sandbox como instrumento de inovação na governança pública, com potencial impacto sobre os rumos da regulação da inteligência artificial no âmbito do Estado brasileiro.

Destaca-se, por fim, que esta é a primeira aplicação da metodologia RESMA (Regulatory Sandbox Maturity Assessment) em uma instituição pública¹ da América Latina. Trata-se de um referencial internacional emergente, que fornece um arcabouço rigoroso para avaliação da capacidade legal, institucional e técnica de órgãos públicos na implementação de mecanismos regulatórios experimentais. Os resultados obtidos por meio desta experiência pioneira não apenas influenciarão a evolução da governança de IA no Brasil, como também contribuirão para o aperfeiçoamento da RESMA enquanto instrumento de formulação e avaliação de políticas públicas em nível global.

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>.



ALINHAMENTO COM POLÍTICAS PÚBLICAS E ESTRATÉGIAS INSTITUCIONAIS

A iniciativa de inteligência artificial (IA) da Advocacia-Geral da União (AGU) encontra-se fortemente alinhada com estratégias nacionais e institucionais que priorizam a transformação digital, a inovação jurídica e a modernização da administração pública. Tal iniciativa reforça os objetivos centrais do Plano de Desenvolvimento Institucional da AGU e de sua Agenda Estratégica de Inovação, os quais valorizam a promoção da segurança jurídica, a eficiência na prestação dos serviços públicos e a adoção responsável de tecnologias emergentes.



No âmbito nacional, o projeto está em consonância com a Estratégia Brasileira de Inteligência Artificial, que incentiva a experimentação segura de sistemas de IA com base em princípios éticos, de transparência e de responsabilidade. Nesse contexto, o sandbox regulatório proporciona um ambiente de risco controlado, propício à formulação de boas práticas na aplicação de funções jurídicas assistidas por IA.

A iniciativa também contribui para os objetivos da Estratégia de Governo Digital, ao fomentar a inovação nos processos jurídicos digitais, ampliar a agilidade institucional e aprimorar o acesso da sociedade aos serviços públicos.

Adicionalmente, o projeto observa de forma rigorosa os princípios estabelecidos pela Lei Geral de Proteção de Dados Pessoais (LGPD), incorporando um sólido arcabouço de governança de dados. Esse arcabouço fundamenta-se em diretrizes como a privacidade desde a concepção (privacy by design), a limitação e adequação da finalidade (purpose limitation) e a auditabilidade contínua.

Por fim, a proposta encontra respaldo na Política Nacional de Modernização do Estado e na Agenda de Melhoria Regulatória, que preveem a utilização de instrumentos regulatórios inovadores e flexíveis como meios para acomodar soluções tecnológicas disruptivas. Nesse sentido, a iniciativa configura-se não apenas como um experimento tecnológico, mas como um protótipo de transformação institucional com base legal, voltado à promoção da legitimidade democrática e ao fortalecimento da confiança pública.

O IMPACTO NA GOVERNANÇA DE IA NO BRASIL

O sandbox regulatório, em estudo, é concebido como projeto-piloto estratégico para apoiar a integração responsável de sistemas de inteligência artificial (IA) no setor público. Seu objetivo central é demonstrar como instrumentos de governança regulatória experimental podem viabilizar a adoção segura, legítima e socialmente aceitável da IA em domínios tradicionalmente considerados sensíveis ou de alto risco, a exemplo da proteção social, dos serviços jurídicos e do acesso à justiça.

Em lugar de adotar uma abordagem restritiva ou proibitiva quanto ao uso da IA nessas áreas, o sandbox propõe a identificação, avaliação e mitigação dos riscos envolvidos. O ambiente experimental permitirá o teste de estratégias voltadas à conciliação entre inovação tecnológica e observância das garantias constitucionais e legais, com especial atenção a desafios como o viés algorítmico, a explicabilidade das decisões automatizadas e o desenho de mecanismos adequados de supervisão humana — aspectos que ganham ainda mais relevância diante da crescente complexidade e autonomia dos sistemas de IA.

Destaca-se, nesse contexto, o avanço da chamada “IA agente”, sistemas dotados da capacidade de perseguir objetivos de médio e longo prazo por meio de ações autônomas, mesmo na ausência de programação específica para todos os comportamentos possíveis. O grau de agenticidade de tais sistemas pode ser compreendido como sua aptidão para alcançar metas complexas de forma flexível, adaptativa e eficaz em ambientes dinâmicos, muitas vezes com supervisão humana limitada².

A compreensão das implicações da IA para a administração pública é particularmente relevante, sobretudo no que se refere à redefinição do papel dos agentes públicos enquanto usuários ou supervisores desses sistemas. De modo geral, o usuário de um sistema de IA é aquele que o ativa, estabelece seus objetivos operacionais e mantém determinado grau de supervisão sobre seus resultados e interações. No curso de sua operação, tais sistemas podem interagir com terceiros, sejam pessoas ou serviços digitais, a depender da natureza e complexidade das tarefas que lhes são atribuídas.

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>.

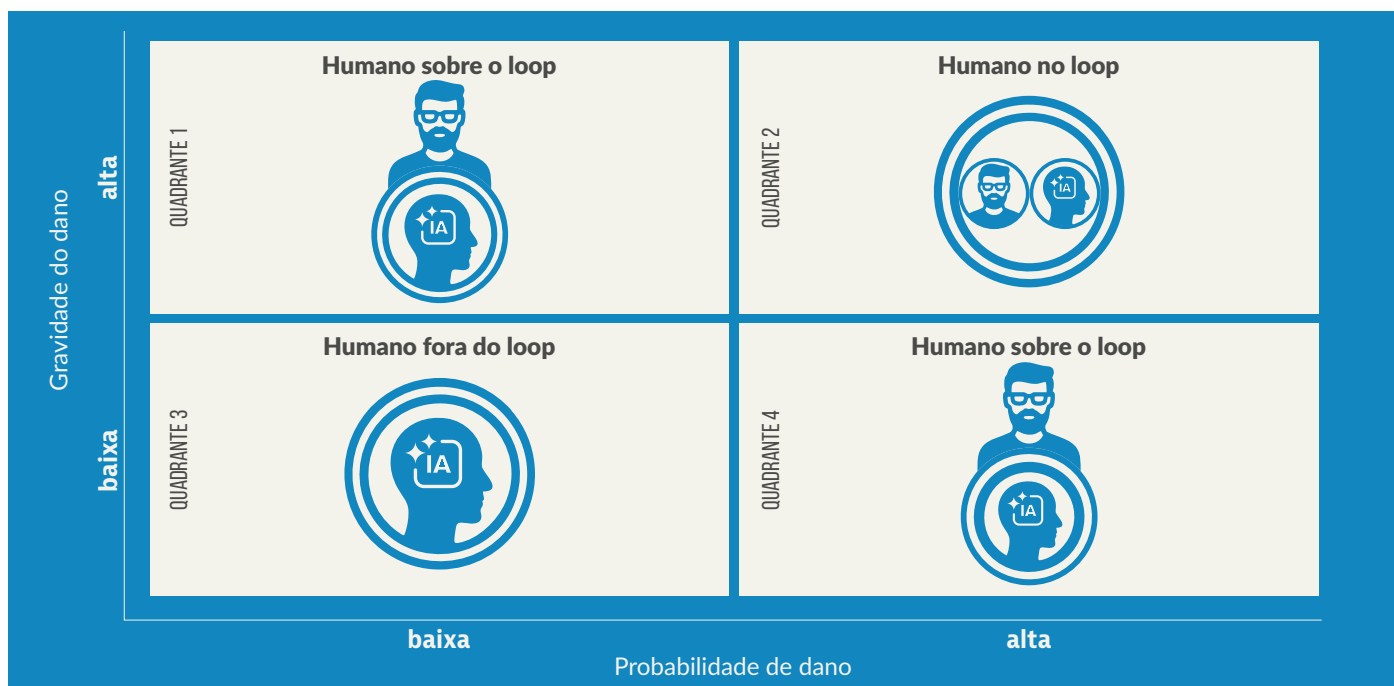


Na condição de usuários ou implementadores de tecnologias baseadas em inteligência artificial (IA), os agentes públicos devem ser capazes de avaliar criticamente a adequação de um sistema à função pretendida. Essa avaliação envolve a verificação da capacidade do sistema de operar com confiabilidade nas condições reais de uso. Em situações em que a confiabilidade absoluta não seja requisito essencial, em virtude da natureza de baixo risco da tarefa, aspectos como o desenho da interface e a orientação ao usuário assumem especial relevância, por contribuírem para a transparência e o alinhamento de expectativas quanto ao desempenho do sistema.

Essas considerações evidenciam a urgência de se estabelecer metodologias robustas para avaliação de sistemas de IA do tipo agente, capazes de antecipar e mitigar possíveis falhas³. O sandbox regulatório, nesse contexto, constitui uma oportunidade estratégica para a geração de evidências empíricas que possam subsidiar avaliações mais amplas e fundamentadas desses sistemas, sobretudo em sua aplicação a funções críticas no âmbito da administração pública brasileira.

Um dos focos centrais do projeto será a definição precisa do papel dos agentes públicos nas decisões assistidas por IA, o que inclui o estabelecimento de parâmetros claros acerca da margem de discricionariedade permitida, das responsabilidades de supervisão e dos correspondentes mecanismos de responsabilização.

Por meio da realização de testes iterativos, a AGU explorará diferentes modelos de supervisão com presença humana (human-in-the-loop), avaliando suas implicações para a legitimidade administrativa e a confiança da sociedade nos processos decisórios mediados por tecnologia. Um modelo internacional de referência particularmente relevante é o adotado por Singapura, cuja abordagem tem despertado expressiva atenção global. Nesse modelo, o grau de supervisão humana é calibrado em função da gravidade potencial dos danos e da probabilidade de sua ocorrência, permitindo o ajuste proporcional do nível de intervenção humana⁴.



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>.



Assim, para atividades de baixo risco, em que tanto a probabilidade quanto o impacto de eventuais danos são mínimos a revisão humana pode ser dispensável. Em contrapartida, em cenários de alto risco, exige-se a atuação ativa e rigorosa de supervisores humanos, assegurando o controle necessário para preservar a integridade do processo decisório e os direitos dos cidadãos.

No contexto brasileiro, o objetivo é adaptar e aperfeiçoar tais conceitos em instituições como a Advocacia-Geral da União (AGU), com especial atenção à definição dos parâmetros que caracterizam o dano, sua natureza, gradações de gravidade e possíveis repercussões para a tomada de decisão pública. O sandbox regulatório proposto oferecerá um ambiente estruturado para o desenvolvimento e teste desses critérios avaliativos, contribuindo de forma concreta para a adoção segura, transparente e juridicamente embasada de sistemas autônomos em áreas sensíveis, como a prestação de serviços públicos, o engajamento cívico e até mesmo os processos judiciais.

As percepções geradas a partir desse experimento regulatório serão fundamentais para a construção de um marco ampliado de governança da inteligência artificial no Brasil, especialmente no que se refere à definição de padrões de supervisão humana apropriados a diferentes graus de risco.

Adicionalmente, espera-se que o sandbox forneça subsídios concretos e tecnicamente qualificados a iniciativas legislativas em curso, como o Projeto de Lei nº 2.338/2023, atualmente em debate no Congresso Nacional. Referido projeto estabelece diretrizes rigorosas de supervisão humana para sistemas de IA classificados como de alto risco, com exigências específicas, tais como:

- que os usuários compreendam as capacidades e limitações do sistema, mantendo controle operacional efetivo;
- que estejam cientes dos riscos associados ao viés de automação e à dependência excessiva dos resultados gerados por sistemas de IA;
- que possuam condições de interpretar adequadamente os resultados à luz da arquitetura do sistema e das ferramentas interpretativas disponíveis.

Ao ancorar a regulação da IA em práticas experimentais concretas e no enfrentamento de desafios reais, esta iniciativa reafirma um princípio basilar da inovação democrática: os marcos regulatórios devem evoluir de maneira responsiva e coordenada com o avanço tecnológico, evitando abordagens estáticas ou dissociadas da realidade institucional.



OBJETIVOS DO SANDBOX



O principal objetivo do sandbox é testar a funcionalidade do sistema de IA em um ambiente real, porém controlado, permitindo à AGU:



Analisar implicações regulatórias, legais e éticas de tarefas jurídicas assistidas por IA;



Avaliar a viabilidade operacional da otimização de fluxos de trabalho internos;



Reforçar a responsabilidade institucional e previsibilidade jurídica;



Gerar evidências empíricas que embasem futuras regulações e estratégias de transformação digital.

IMPLEMENTAÇÃO PILOTO: INDEFERIMENTOS DE BENEFÍCIOS DO INSS



O primeiro caso de aplicação do sandbox envolve disputas administrativas decorrentes do indeferimento de benefícios pelo INSS. O sistema de IA irá:



Detectar padrões entre os casos negados e destacar situações potencialmente elegíveis;



Gerar recomendações e indicadores de risco jurídico para revisão interna;



Apoiar a negociação e redação de acordos extrajudiciais;



Automatizar a geração de documentos jurídicos internos;



Servir como protótipo para futuras aplicações em prevenção e resolução de riscos jurídicos.

ESTRUTURA DE GOVERNANÇA

O sandbox será coordenado por um Comitê de Supervisão, composto por:



O LABORATÓRIO DE INOVAÇÃO (LABORI/ AGU);



A SECRETARIA DE GOVERNANÇA E GESTÃO ESTRATÉGICA (SGE/AGU).

Este comitê supervisionará a validação de hipóteses, autorizará flexibilizações procedimentais e decidirá sobre a ampliação, revisão ou encerramento das soluções testadas. Outras unidades da AGU ou especialistas externos (jurídicos, acadêmicos ou técnicos) poderão ser convidados conforme necessário.

DURAÇÃO E METODOLOGIA

O sandbox terá duração de 12 meses, divididos em quatro etapas:

ETAPA 1. PLANEJAMENTO INTERNO (2–3 MESES):

Definição de objetivos, hipóteses, protocolos de risco e mapeamentos legais.

ETAPA 2. TREINAMENTO E PREPARAÇÃO (1–2 MESES):

Capacitação das equipes; implementação de protocolos do sandbox.

ETAPA 3. EXPERIMENTAÇÃO E TESTES (6–7 MESES):

Execução dos testes com monitoramento contínuo e ajustes.

ETAPA 4. AVALIAÇÃO E APRENDIZADO (2–3 MESES):

Avaliação abrangente de desempenho e recomendações de diretrizes a serem formuladas.



HIPÓTESES REGULATÓRIAS

As hipóteses a serem testadas incluem:



Analisar implicações regulatórias, legais e éticas de tarefas jurídicas assistidas por IA;



Aceitação institucional da redação jurídica assistida por IA;



Legalidade do uso de dados para treinamento e validação de modelos;



Papel da IA na promoção de acordos extrajudiciais;



Explicabilidade e auditabilidade dos resultados produzidos pela IA



FLEXIBILIDADES REGULATÓRIAS



Durante o período do sandbox, poderão ser aplicadas exceções regulatórias temporárias, como:



Uso de IA para redação de documentos jurídicos, suspendendo exigências de elaboração exclusivamente manual;



Ajuste de prazos internos para permitir refinamento iterativo;



Uso de ferramentas experimentais de autenticação em conformidade com as políticas de segurança da AGU;



Uso de dados pseudonimizados ou anonimizados, conforme os protocolos da Política de Segurança da Informação da Advocacia-Geral da União (POSIN-AGU);



Ajustes controlados no acesso ao sistema e rastreabilidade, com supervisão rigorosa;



Regime regulatório provisório aplicável apenas no ambiente do sandbox;

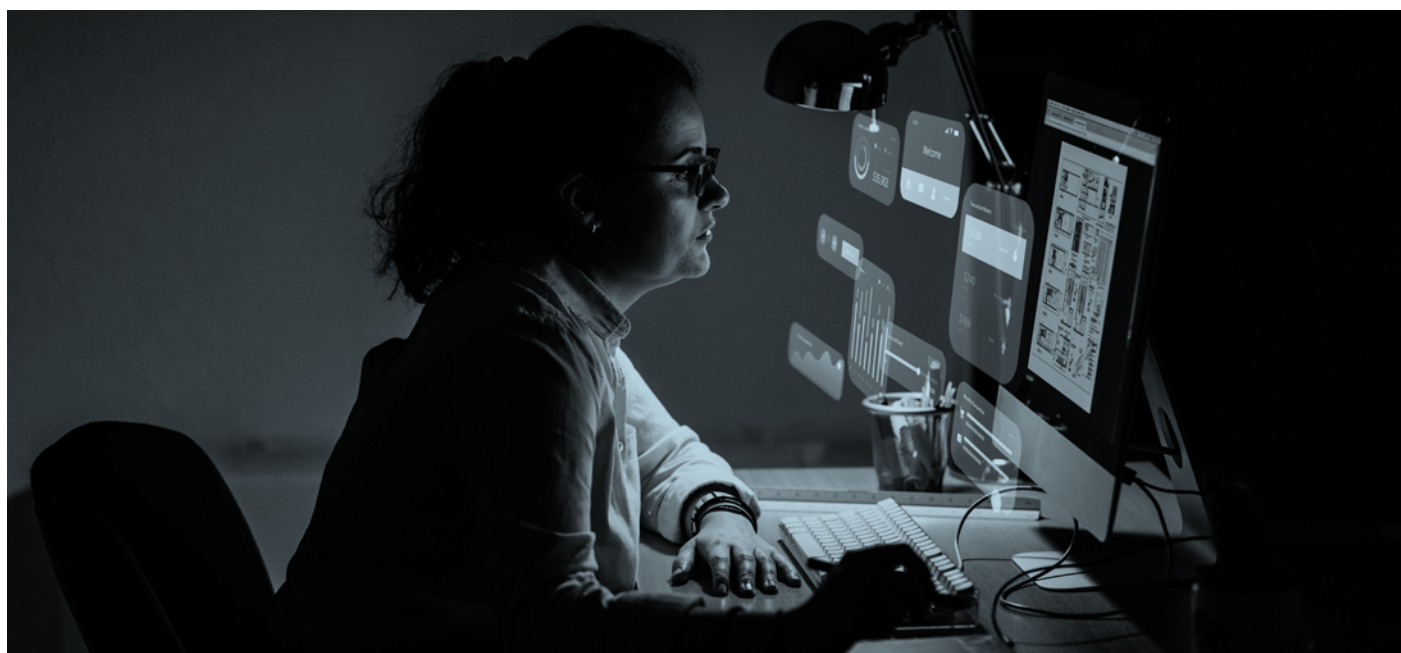


Protocolos de documentação e auditoria adaptados para garantir responsabilidade e reversibilidade.

Todas as flexibilidades respeitarão os princípios fundamentais da legalidade, devido processo legal, proteção de dados pessoais, transparência institucional e possibilidade de revogação em caso de risco.

MONITORAMENTO E AVALIAÇÃO

O monitoramento robusto garantirá o cumprimento contínuo das normas legais, éticas e de proteção de dados. Mecanismos de avaliação em tempo real acompanharão os resultados do sistema, detectarão anomalias e permitirão correções rápidas. Os critérios de avaliação enfatizarão a integridade institucional, o desempenho técnico e a confiança dos stakeholders.



ENTREGAS FINAIS



O Comitê de Supervisão emitirá um Relatório Final de Avaliação, que incluirá:



Avaliação técnica e jurídica de desempenho;



Síntese de aprendizados regulatórios e jurídicos;



Recomendações para adoção, revisão ou encerramento institucional da tecnologia;



Minutas de instrumentos normativos, conforme necessário para institucionalização.

CONSIDERAÇÕES FINAIS



Esta iniciativa foi viabilizada com o apoio das instituições GIZ e CEPAL, marcando uma aplicação pioneira da metodologia RESMA. Dentre as principais contribuições do projeto estão:



Capacidade de definir previamente as condições para experimentação regulatória estruturada;



Precisão na identificação e teste de prioridades regulatórias;



Impacto significativo no panorama brasileiro de governança de IA e o valor dos resultados documentados.

Como uma das primeiras implementações da RESMA a ser documentada, os aprendizados deste sandbox oferecem insights relevantes para o futuro da experimentação regulatória em IA. Os resultados indicam melhorias na eficiência, redução de prazos e uso mais eficaz de recursos — fortalecendo a continuidade do desenvolvimento de mecanismos de governança experimental no setor público.



Labori
Laboratório de Inovação da AGU

