

CONTROLADORIA-GERAL DA UNIÃO

Secretaria-Executiva

Diretoria de Planejamento, Inovação e Sustentabilidade



GUIA DE USO RESPONSÁVEL DE FERRAMENTAS DE *IA GENERATIVA*

Brasília - Março 2025

CONTROLADORIA-GERAL DA UNIÃO – CGU

Setor de Autarquias Sul (SAUS), Quadra 5, Bloco A

Ed. Multibrasil, Brasília/DF - CEP: 70.070-050

cgu@cgu.gov.br

VINÍCIUS MARQUES DE CARVALHO

Ministro da Controladoria-Geral da União

EVELINE MARTINS BRITO

Secretária-Executiva

RONALD DA SILVA BALBE

Secretário Federal de Controle Interno

RICARDO WAGNER DE ARAÚJO

Corregedor-Geral da União

ARIANA FRANCES CARVALHO DE SOUZA

Ouvidora-Geral da União

MARCELO PONTES VIANNA

Secretário de Integridade Privada

LIVIA OLIVEIRA SOBOTA

Secretária de Integridade Pública

ANA TÚLIA DE MACEDO

Secretária Nacional de Acesso à Informação

Sumário

1. Introdução	4
2. Glossário de Termos	4
3. Diretrizes no uso de ferramentas de IA Generativas	6
4. Usos para IA generativas	7
5. Recomendações no uso de IA generativa	8
6. Como avaliar uma resposta da IA Generativa	9
7. Recomendações de segurança no uso de ferramentas públicas de IA Generativa	10
8. Considerações finais	11
9. Referências	11

1. Introdução

- 1.1. Este guia esclarece conceitos e estabelece orientações para o uso responsável de ferramentas de Inteligência Artificial Generativa na Controladoria-Geral da União (CGU).
- 1.2. As diretrizes e orientações aqui contidas aplicam-se a servidores, estagiários e terceirizados no uso de aplicativos de inteligência artificial generativa, abrangendo tanto modelos e soluções internos quanto modelos de terceiros e aplicativos disponíveis publicamente.
- 1.3. A adoção de ferramentas de IA generativa na CGU pode trazer oportunidades significativas, especialmente em termos de produtividade na administração pública. Com sua capacidade de processar grandes volumes de dados rapidamente e gerar insights, as ferramentas de IA podem ajudar os usuários da CGU, promovendo uma gestão mais eficiente, ágil, responsiva e capaz de atender às necessidades dos cidadãos.

2. Glossário de Termos

- 2.1. LIA: é a plataforma de inteligência artificial generativa disponibilizada pela Diretoria de Tecnologia da Informação.
- 2.2. CGU-Insight: é uma aplicação web que emprega modelos de Inteligência Artificial para auxiliar na condução de auditorias, oferecendo um conjunto de funcionalidades focadas na análise de documentos e geração de insights relevantes para os trabalhos de auditoria.
- 2.3. AuditPesquisa: é um sistema inteligente, desenvolvido para auxiliar na resposta de perguntas relacionadas aos relatórios de auditoria da CGU. Utilizando a técnica chamada RAG, o sistema combina a busca de informações relevantes com a criação de respostas precisas. Integrado a um grande modelo de linguagem (LLM) como o ChatGPT, o AuditPesquisa oferece respostas mais precisas e atualizadas, buscando informações diretamente dos relatórios de auditoria.
- 2.4. Eva: é uma IA baseada em GPT que orienta os usuários do Sistema ePAD na condução dos procedimentos disciplinares. Ela oferece conhecimento detalhado sobre a matéria correcional e proporciona acesso rápido a orientações sobre como apurar e investigar condutas de servidores públicos. A base de conhecimento da Eva inclui legislações, orientações e protocolos específicos, auxiliando os membros das comissões a realizarem suas tarefas com maior precisão e segurança.
- 2.5. RAG (Retrieval-Augmented Generation): no contexto de aprendizado de máquina e processamento de linguagem natural, RAG refere-se a uma arquitetura de modelo que combina abordagens baseadas em recuperação e geração para melhorar a qualidade do texto gerado, incorporando informações relevantes de fontes externas.
- 2.6. Modelo de Linguagem: Um modelo de linguagem é um sistema de inteligência artificial capaz de reconhecer, compreender ou gerar texto em linguagem humana baseado em exemplos de treinamento.

- 2.7. Modelos de Linguagem *Open Source*: modelos de linguagem cujo código-fonte é disponibilizado de forma aberta ao público, de modo que qualquer pessoa pode acessar, utilizar, modificar e distribuir esse código-fonte de forma livre.
- 2.8. Inteligência Artificial Generativa: é um tipo de IA que pode criar conteúdo e ideias, incluindo conversas, histórias, imagens, vídeos e músicas.
- 2.9. Plataformas Públicas de IA generativa: Soluções de IA generativa, fornecidas por terceiros. São exemplos desses tipos de solução:
- 2.9.1. ChatGPT, Copilot, LLaMA e Gemini, que podem fornecer respostas em linguagem natural sobre basicamente qualquer tipo de assunto a partir de um prompt inicial;
 - 2.9.2. GitHub Copilot e FauxPilot, que podem produzir códigos de programação a partir de prompts de texto;
 - 2.9.3. DALL-E, Midjourney e Stable Diffusion, que podem produzir imagens a partir de textos ou mesmo de outras imagens.
- 2.10. Prompt: Comando de texto dado a um modelo de linguagem de IA para gerar uma resposta ou realizar uma tarefa específica. A qualidade e precisão da resposta podem variar significativamente com a formulação do prompt.
- 2.11. Alucinação: termo usado na IA generativa para descrever respostas fictícias, confiantes e convincentes, que podem ser erroneamente aceitas devido a algum viés e podendo escapar a uma revisão superficial por quem não conhece profundamente do assunto.
- 2.12. Racismo algorítmico: refere-se ao viés discriminatório que surge quando algoritmos, incluindo aqueles usados para treinar modelos de IA, perpetuam ou amplificam preconceitos raciais. Isso se manifesta quando esses modelos geram conteúdos que reforçam estereótipos raciais ou que respondem de forma diferenciada com base em características raciais implícitas nos dados
- 2.13. Viés de resultado: também chamado de viés de aprendizado de máquina ou viés de algoritmo, refere-se à ocorrência de resultados tendenciosos devido a vieses humanos que distorcem os dados de treinamento originais ou o algoritmo de IA, levando a resultados distorcidos e potencialmente prejudiciais.
- 2.14. Persona: é uma representação fictícia de um usuário ou cliente ideal, baseada em dados reais sobre o comportamento, necessidades, motivações e desafios de um grupo-alvo. São exemplos de personas: servidor público, auditor, diretor, usuário padrão, etc.
- 2.15. Deepfake: é o nome dado a utilização da inteligência artificial para criar ou manipular vídeos, áudios ou imagens de forma realista, fazendo parecer que uma pessoa está dizendo ou fazendo algo que na realidade não fez. Os deepfakes podem alterar rostos em vídeos, modificar expressões faciais e até imitar vozes, criando conteúdo que pode ser extremamente convincente.
- 2.16. Opt-out: refere-se ao processo pelo qual um indivíduo ou organização decide não participar de uma atividade específica, como o recebimento de

comunicações ou a coleta de dados, geralmente após ser informado sobre essa possibilidade. Essa prática é importante para garantir a privacidade e o controle do usuário sobre suas informações pessoais, permitindo que ele escolha o que deseja ou não receber.

- 2.17. *Phishing*: é uma técnica de fraude digital na qual um atacante se disfarça como uma entidade confiável para enganar indivíduos e obter informações sensíveis, como senhas ou dados financeiros. Isso é frequentemente realizado por meio de e-mails, mensagens de texto ou sites falsificados, levando os usuários a fornecer dados pessoais involuntariamente, o que pode resultar em roubo de identidade ou outras formas de prejuízo.
- 2.18. *Malware*: é uma abreviação de "*malicious software*" e refere-se a qualquer software projetado para causar danos, explorar ou acessar de forma não autorizada dispositivos e sistemas. Isso inclui vírus, *worms*, *trojans* e *ransomware*, entre outros, que podem comprometer a segurança das informações, danificar sistemas e causar perdas financeiras, sendo uma preocupação constante na segurança cibernética.

3. Diretrizes no uso de ferramentas de IA Generativas

- 3.1. Prevalência da pessoa humana nas decisões: o usuário deve manter o controle humano sobre o processo criativo. A IA generativa deve ser uma ferramenta para auxiliar e ampliar a criatividade, não para substituí-la.
- 3.2. Responsabilidade do usuário sobre as informações utilizadas: o usuário é o responsável pelo uso de informações obtidas a partir de ferramentas de IA generativa, cabendo a este a revisão crítica de todos os dados e informações. Dessa forma, o usuário deve ser capaz de explicar, justificar e assumir a responsabilidade por suas orientações e decisões.
- 3.3. Privacidade e proteção de dados: os dados e informações protegidos legalmente não devem ser utilizados em ferramentas públicas de IA generativa.
- 3.4. Alinhamento aos princípios constitucionais: o uso das ferramentas de IA generativa devem observar os princípios constitucionais da legalidade, impessoalidade, moralidade, publicidade e eficiência.
- 3.5. Produtividade: as ferramentas de IA Generativa na CGU têm o foco de aumentar a eficiência dos usuários e serviços. Dessa forma, o uso deve focar em atividades que possam ser otimizadas pelo uso de recursos computacionais.
- 3.6. Transparência: é importante informar que uma IA generativa está sendo usada para gerar conteúdo, garantindo que os usuários saibam a origem das informações.
- 3.7. Treinamento e conscientização: deve-se educar as equipes sobre as capacidades e limitações da IA generativa, promovendo uma compreensão clara dos riscos, tais quais as alucinações e os vieses que as respostas podem apresentar, e benefícios, com o fim de assegurar melhor utilização e aplicação das ferramentas.

4. Usos para IA generativas

- 4.1. As ferramentas de IA generativa podem transformar a forma de trabalho dos usuários. Os usos mais comuns para IA generativa são:
 - 4.1.1. Elaboração de Documentos: ajudar na redação de relatórios, comunicados e outros documentos oficiais, garantindo clareza e consistência.
 - 4.1.2. Revisão e Tradução: revisar documentos para correção gramatical e estilística, além de traduzir textos entre diferentes idiomas.
 - 4.1.3. Resumo de textos e documentos: transformar grandes textos em resumos claros e diretos de forma rápida e precisa. Ela garante que os resumos mantenham a essência do texto original, facilitando a tomada de decisões e o acesso rápido às informações.
 - 4.1.4. Classificação de dados: Classificar os dados existentes em documentos em relação a uma determinada categoria.
 - 4.1.5. Responder sobre documentos: ler e processar grandes volumes de texto em questão de segundos, extraíndo informações relevantes e resumindo os pontos principais. Faça perguntas sobre qualquer documento e a IA generativa fornecerá respostas precisas e informativas, mesmo que sejam complexas ou abertas.
 - 4.1.6. Consulta a Normas e Regulamentos: facilitar o acesso a informações sobre legislações e normas, ajudando na interpretação de leis e regulamentos.
 - 4.1.7. Processamento de Dados em Massa: analisar grandes volumes de dados para identificar padrões e anomalias que possam indicar fraudes ou irregularidades.
 - 4.1.8. Criação de bases de dados para testes: elaborar bases de dados para a realização de testes em sistemas. Essas bases podem ser geradas nas mais diversas linguagens de programação, facilitando o processo de upload de dados.
 - 4.1.9. Geração de imagens, áudios ou vídeos: é possível criar conteúdos dos mais diversos tipos, facilitando o processo criativo. Esse recurso requer cautela e responsabilidade do usuário, uma vez que é possível simular a fala e até mesmo criar vídeos falsos de pessoas reais (*deepfake*), com um nível de similaridade e realismo em que é quase impossível a uma pessoa média distinguir se aquilo é real ou não.
 - 4.1.10. Transcrição e tradução de arquivos de áudio e vídeo em tempo real: a ferramenta de IA generativa pode ouvir e transcrever conversas ou reuniões ou mesmo traduzir o conteúdo, otimizando o processo de geração de atas de reunião.
- 4.2. Entretanto, dada algumas limitações dos modelos atuais, não é recomendado o uso de ferramentas de IA generativa para funções como:
 - 4.2.1. Predição/previsão de cenários e dados: por causa dos riscos de erros ou

incertezas em estimativas futuras.

4.2.2. Planejamento e otimização: pois decisões estratégicas exigem análise humana profunda e ponderada.

4.2.3. Inteligência decisória: em que a complexidade e o contexto específico das decisões precisam ser compreendidos por especialistas.

5. Recomendações no uso de IA generativa

- 5.1. Ao considerar o uso de ferramentas de IA generativas em seu trabalho, o usuário deve avaliar os possíveis benefícios e riscos de cada caso de uso e tomar as medidas adequadas para mitigá-los.
- 5.2. Priorize o uso de ferramentas de IA generativa disponibilizadas pela Controladoria-Geral da União, pois estas possuem recursos de segurança que mitigam riscos de vazamentos de dados e informações sensíveis protegidas legalmente.
- 5.3. Analise os resultados dos aplicativos de IA generativa para garantir que eles atendam aos padrões da CGU quanto aos princípios que orientam a atuação da Administração Pública.
- 5.4. Avalie o conteúdo gerado por IA generativa para garantir que não discrimine indivíduos com base em raça, cor, religião, sexo, nacionalidade, idade, deficiência, estado civil, afiliação política ou orientação sexual.
- 5.5. Sempre avalie e revise criteriosamente o conteúdo gerado por IA Generativa, ainda que o sistema pareça confiável, de modo a garantir o uso de respostas precisas e apropriadas para o fim a que se destina.
- 5.6. Considere criteriosamente sua capacidade de identificar eventuais imprecisões de conteúdos gerados antes de usar a IA Generativa. Evite o seu uso se considerar que não conseguirá validar o conteúdo gerado.
- 5.7. Não implemente nem use código de programação gerado por IA generativa nos sistemas da instituição sem revisão minuciosa por especialista de TI.
- 5.8. Ao desenvolver ferramentas ou processos que incorporem IA generativa, certifique-se de continuar a monitorar o desempenho em relação à finalidade pretendida. Os modelos de IA generativa são atualizados de tempos em tempos e podem aprender com os dados com os quais interagem. Isso pode afetar o desempenho e a confiabilidade do resultado.
- 5.9. Não use em nenhum material institucional resultados que contenham material que se suspeite possam estar sob proteção de direitos autorais.
- 5.10. Use o recurso "*opt-out*", quando possível, para tentar reduzir a utilização dos prompts no treinamento e desenvolvimento de um sistema de IA.
- 5.11. Comunique claramente quando e como você está usando a IA nas interações com o público.
- 5.12. Informe aos usuários quando mensagens de texto, áudio ou visuais dirigidas a eles forem geradas pela IA. As pessoas podem não saber que estão

interagindo com um sistema de IA, ou podem presumir erroneamente que a IA está sendo usada.

5.13. Para obter melhores resultados:

5.13.1. Inclua detalhes em sua consulta para obter respostas mais relevantes;

5.13.2. Peça ao modelo para adotar uma persona;

5.13.3. Instrua o modelo a responder usando uma citação ou texto de referência;

5.13.4. Defina a quantidade de texto ou tamanho que será gerado como resultado do prompt;

5.13.5. Divida suas consultas em etapas menores e específicas, conduzindo uma sequência lógica de perguntas. Essa abordagem ajuda a manter um maior controle sobre as respostas geradas, melhora a qualidade do resultado e reduz o risco de exposição de informações sensíveis.

5.13.6. Pergunte ao modelo se ele deixou passar alguma informação nas passagens anteriores. Solicite novas respostas até obter um resultado adequado.

6. Como avaliar uma resposta da IA Generativa

6.1. O uso de ferramentas de IA está sujeito a riscos, sendo os principais os fenômenos da alucinação, do viés de resultado e da desatualização dos dados de treinamento.

6.2. A alucinação em ferramentas de IA generativa refere-se ao fenômeno em que o modelo de IA produz informações que parecem plausíveis, mas são incorretas, imprecisas ou completamente inventadas.

6.3. Já o viés de resultado refere-se à reprodução de vieses presentes nos dados de treinamento, o que pode levar a resultados enganosos ou injustos. O viés nos resultados pode afetar desproporcionalmente alguns grupos, como pessoas pretas, pessoas com deficiência, comunidades LGBTQ+, dentre outros.

6.4. A desatualização dos dados de treinamento em ferramentas de IA generativa refere-se à falta de atualização ou renovação das informações que alimentam o modelo, resultando em um conhecimento defasado ou incompleto. Nesse contexto, a IA generativa está suscetível a fornecer respostas imprecisas, inadequadas ou baseadas em informações obsoletas, comprometendo a utilidade e a confiabilidade da ferramenta, especialmente em contextos em que a informação precisa ser atual e precisa.

6.5. A fim de evitar essas questões, bem como aperfeiçoar a qualidade do resultado gerado por uma IA generativa, recomenda-se ao usuário:

6.5.1. Revisar a pergunta original: certifique-se de entender completamente a pergunta ou o tópico abordado. Leia a pergunta várias vezes. Assegure-se que todas as palavras-chave e termos técnicos são compreensíveis.

6.5.2. Fatos e Dados: verifique se a resposta contém fatos e dados precisos. Uma boa resposta deve incluir referências ou citar fontes quando possível. O usuário pode usar fontes confiáveis, como enciclopédias, artigos acadêmicos,

ou sites oficiais, para comparar a informação e validar a credibilidade da resposta.

- 6.5.3. Recentidade: fique atento acerca do período sobre o qual se busca respostas, uma vez que podem existir dados mais recentes do que aqueles que a IA generativa tem disponível para elaborar a resposta. Busque conhecer a data da última atualização da base de dados do modelo em uso e complemente a resposta com uma pesquisa independente, caso o período de atualização seja anterior à necessidade.
- 6.5.4. Coerência: verifique se a resposta é coerente internamente e se as afirmações fazem sentido entre si. Revise a resposta para garantir que todas as partes se complementam e que não há contradições ou incoerências.
- 6.5.5. Abrangência: avalie se a resposta cobre todos os aspectos da pergunta original.
- 6.5.6. Profundidade: verifique se a resposta vai além de uma explicação superficial e oferece detalhes suficientes. Forneça explicações detalhadas, exemplos e contextos quando formular a pergunta inicial ou solicitar um complemento ao resultado recebido.
- 6.5.7. Linguagem Simples: a resposta deve ser escrita em uma linguagem clara e acessível, sem jargões desnecessários e termos técnicos complexos sem explicação.
- 6.5.8. Estrutura: a organização da resposta deve ser lógica e fácil de seguir.
- 6.5.9. Objetividade: verifique se a resposta é imparcial e não contém viés evidente. Respostas imparciais apresentam diferentes pontos de vista quando aplicável. A IA generativa deve oferecer uma visão neutra sem tomar partido em questões subjetivas.
- 6.5.10. Pertinência: certifique-se de que cada parte da resposta é relevante para a pergunta e contribui para a compreensão do tema. Quando não delimitado na pergunta, as ferramentas de IA generativa tendem a ser prolixas, sendo necessário revisar a resposta para eliminar partes irrelevantes ou que desviam do tópico central.
- 6.5.11. Ajustes: se possível, ajuste a pergunta ou solicite uma nova resposta da IA com base no feedback fornecido para ver melhorias na qualidade.

7. Recomendações de segurança no uso de ferramentas públicas de IA Generativa

- 7.1. Evite usar e-mails, números de telefone ou qualquer outra credencial institucional ao criar contas em ferramentas de IA generativa disponíveis publicamente. Essa medida ajuda a proteger a privacidade e a segurança dos dados da instituição.
- 7.2. Não insira informações internas da instituição que estejam protegidas por algum nível de sigilo em ferramentas de IA generativa que não seja uma solução disponibilizada pela CGU.
- 7.3. Não insira informações que permitam que as ferramentas públicas de IA

generativa extrapolem informações confidenciais ou classificadas com base na agregação do conteúdo que o próprio usuário inseriu ao longo do tempo.

- 7.4. Seja cauteloso ao clicar em links fornecidos ou gerados por ferramentas públicas de IA generativa, pois eles podem levar a sites de *phishing* ou a downloads de *malware*. Prefira clicar em links de fontes confiáveis e verificados. Se o link for solicitado explicitamente por você e provenha de uma ferramenta ou fonte que você conhece, avalie cuidadosamente antes de acessá-lo.
- 7.5. Não distribua ou clique em link fornecido ou gerado por ferramentas públicas de IA generativa. Esses links podem levar a sites de *phishing* ou a downloads de *malware*. Clique apenas em links de fontes confiáveis.
- 7.6. Não insira informações pessoais de servidores, cidadãos ou outros terceiros em nenhuma ferramenta de IA generativa que não seja uma solução disponibilizada pela CGU. Dados abertos ao público na internet podem ser utilizados sem restrição.

8. Considerações finais

- 8.1. A adoção das boas práticas de IA generativa é fundamental para garantir a segurança, a ética e a eficiência no uso dessa tecnologia nas diferentes unidades da Controladoria-Geral da União. Neste sentido, este documento busca fornecer diretrizes para que os usuários possam operar com responsabilidade, respeitando a privacidade dos dados tratados e a qualidade dos serviços prestados.
- 8.2. Em caso dúvidas ou sugestões sobre o conteúdo apresentado, entre em contato por meio do e-mail: suporte.ia@cgu.gov.br.
- 8.3. Em situações de vazamento de dados ou qualquer suspeita de comprometimento de informações sensíveis, pedimos que entre em contato imediatamente pelo e-mail abuse@cgu.gov.br, para que as medidas cabíveis possam ser tomadas com urgência e de maneira adequada.

9. Referências

AUSTRALIAN GOVERNMENT ARCHITECTURE. Interim guidance on government use of public generative AI tools. November 2023. Disponível em: <https://architecture.digital.gov.au/guidance-generative-ai>. Acesso em: 30 jul. 2024.

BRASIL. Projeto de Lei nº 2338, de 2023. Autoria: Senador Rodrigo Pacheco (PSD/MG).

CANADÁ. Guide on the use of generative artificial intelligence. Gouvernement du Canada. Disponível em: <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/guide-use-generative-ai.html>. Acesso em: 30 jul. 2024.

CONSELHO NACIONAL DE JUSTIÇA. Resolução nº 332, de 21 de agosto de 2020. Brasília: CNJ, 2020.

SAN FRANCISCO. San Francisco generative AI guidelines. December 11, 2023. Disponível em: <https://www.sf.gov/reports/december-2023/san-francisco-generative-ai-guidelines>. Acesso em: 30 jul. 2024.

SILVA, Fernanda dos Santos Rodrigues; SILVA, Tarcízio (orgs.). Inteligência artificial e discriminação racial no Brasil: questões principais e recomendações. Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2024. Disponível em: <https://irisbh.com.br/publicacoes/inteligencia-artificial-e-discriminacao-racial-no-brasil/>. Acesso em: 09 out. 2024.

TRIBUNAL DE CONTAS DA UNIÃO. Guia de uso de inteligência artificial generativa no Tribunal de Contas da União (TCU). Brasília: TCU, 2024.