

ADMINISTRAÇÃO
PÚBLICA



GOVERNANÇA
USO ÉTICO E RESPONSÁVEL



SEGURANÇA DA INFORMAÇÃO
CONFIDENCIALIDADE
INTEGRIDADE
DISPONIBILIDADE



SUPERVISÃO HUMANA
DECISÃO E RESPONSABILIDADE
DO AGENTE PÚBLICO



CONFORMIDADE LEGAL
LGPD
NORMAS INTERNAS
LEGISLAÇÃO APLICÁVEL



TRANSPARÊNCIA
RASTRABILIDADE E
PRESTAÇÃO DE CONTAS



ANÁLISE DE RISCOS
IDENTIFICAR
AVALIAR
MITIGAR
MONITORAR



DADOS



MODELO DE IA



VALIDAÇÃO HUMANA



DECISÃO
RESPONSÁVEL



Diretrizes Éticas e Análise de Riscos no Uso de Inteligência Artificial

Princípios, responsabilidades e riscos para o uso ético,
seguro e responsável da IA na Administração Pública.

Sumário

1.	Introdução, Objetivos e Fundamentos	3
2.	Princípios e Diretrizes Gerais para o Uso da IA.....	4
2.1.	Centralidade da decisão humana.....	4
2.2.	Responsabilidade pelo resultado	4
2.3.	Transparência no uso da tecnologia	5
2.4.	Segurança da informação e proteção de dados	5
2.5.	Adequação ao interesse público	5
2.6.	Prudência e uso proporcional da tecnologia	6
3.	Riscos Envolvidos na Adoção, Desenvolvimento e Uso de IA	7
3.1.	Confiabilidade das informações e validação do conteúdo	7
3.2.	Proteção de dados e controle da informação	7
3.3.	Adequação ao contexto institucional.....	8
3.4.	Dependência tecnológica	8
3.5.	Transparência, rastreabilidade e capacidade de explicação	8
3.6.	Uso desproporcional ou inadequado da tecnologia	9
3.7.	Modelo de Avaliação de Riscos ao Longo do Ciclo de Uso da IA.....	10
3.7.1.	Extração e consolidação dos dados de entrada	10
3.7.2.	Entrada ou exposição dos dados de entrada	10
3.7.3.	Processamento ou execução	10
3.7.4.	Compilação e organização dos dados de saída	10
3.7.5.	Utilização dos dados e informações de saída	11
3.8.	Modelo de Avaliação de Riscos associado ao desenvolvimento de projetos .	11
4.	Considerações Finais	12
5.	Referências	13

1. Introdução, Objetivos e Fundamentos

A crescente adoção de soluções baseadas em Inteligência Artificial (IA) na Administração Pública impõe não apenas a necessidade de orientação prática aos servidores, mas também a definição de um conjunto estruturado de diretrizes que sustentem o uso seguro, ético e juridicamente adequado dessas tecnologias.

Nesse contexto, este documento de **Diretrizes e Análise de Riscos no Uso de IA** foi concebido como um instrumento norteador ao uso de inteligência artificial no âmbito da Autarquia. Este documento busca oferecer uma visão estruturante, voltada à compreensão dos princípios, responsabilidades e riscos associados ao uso da inteligência artificial no ambiente institucional.

Assim, as diretrizes aqui apresentadas não devem ser interpretadas como regras abstratas, mas como o alicerce conceitual e normativo que dá suporte às condutas práticas recomendadas.

A Inteligência Artificial deve ser compreendida, no âmbito institucional, como uma tecnologia capaz de ampliar a capacidade analítica, reduzir o tempo de execução de tarefas e facilitar o tratamento de informações complexas. Contudo, tais facilidades requerem atenção aos riscos envolvidos.

Esses riscos não se restringem a falhas técnicas, como também abrangem aspectos relacionados à privacidade, à segurança da informação, à transparência, e à confiabilidade das decisões. Por essa razão, o uso da IA exige uma abordagem que combine inovação com prudência, assegurando que o ganho de eficiência não ocorra em detrimento dos direitos dos cidadãos ou da integridade institucional.

Além disso, este documento se alinha às diretrizes mais amplas da Estratégia Brasileira de Inteligência Artificial (EBIA), especialmente no que se refere à promoção do uso ético, responsável e confiável da tecnologia no setor público. Tal alinhamento reforça o compromisso institucional com a construção de um ambiente de inovação que respeite os valores democráticos, a legislação vigente e o interesse público.

Por fim, é importante destacar que a efetividade destas diretrizes depende tanto da existência formal de regras quanto da sua incorporação no cotidiano institucional.

2. Princípios e Diretrizes Gerais para o Uso da IA

A definição de diretrizes para o uso da Inteligência Artificial na Autarquia parte do reconhecimento de que a tecnologia não é neutra nem infalível. Seu uso depende da forma como é incorporada às rotinas de trabalho, das decisões tomadas pelos usuários e dos limites institucionais estabelecidos.

Nesse sentido, torna-se necessário consolidar um conjunto de princípios que nortearão o uso, a adoção e o desenvolvimento de ferramentas de Inteligência Artificial e que funcionam como referência para balizar a conduta do servidor.

De forma integrada, os princípios a seguir oferecem ao servidor um referencial para orientar suas decisões no uso da Inteligência Artificial. Eles funcionam como critérios de julgamento, permitindo adaptar o uso da tecnologia aos diferentes contextos de trabalho sem perder o alinhamento com as responsabilidades públicas.

2.1. Centralidade da decisão humana

A Inteligência Artificial deve sempre ocupar uma posição de apoio, e não de substituição do agente público. Isso significa que:

- o conteúdo produzido por ferramentas de IA não deve ser adotado sem ser revisado;
- decisões não podem ser delegadas à tecnologia, mas podem ser subsidiadas com auxílio dela;
- o servidor permanece como responsável por avaliar a pertinência, a legalidade e os impactos de qualquer resultado gerado.

Desse modo, a IA pode otimizar a capacidade do servidor, mas não assume sua função decisória.

2.2. Responsabilidade pelo resultado

O uso de IA não isenta a responsabilidade o servidor no exercício da função pública. Assim:

- qualquer documento, análise ou manifestação institucional elaborada com apoio de IA continua sendo de responsabilidade integral do servidor;
- erros, omissões ou inconsistências não podem ser atribuídos à ferramenta;
- a validação deve abranger tanto a forma quanto o conteúdo técnico gerado.

Esse princípio reforça que a tecnologia não possui autonomia institucional — ela apenas instrumentaliza a atividade humana.

2.3. Transparência no uso da tecnologia

A utilização de sistemas de IA deve ser comunicada de forma clara às pessoas afetadas, especialmente quando decisões algorítmicas impactam seus direitos.

De modo geral, espera-se que o servidor:

- informe previamente os usuários quando uma decisão for fundamentada ou apoiada por sistemas de IA;
- assegure que os indivíduos possam solicitar explicações sobre as decisões que os afetam e compreender os critérios utilizados;
- garanta o acesso a um responsável capaz de revisar e, se necessário, corrigir a decisão;
- comunique de maneira adequada quando produtos ou serviços forem fornecidos diretamente ou com o auxílio de sistemas de IA.

O uso de IA deve sempre preservar a possibilidade de controle humano, assegurando às pessoas o direito de questionar, compreender e contestar decisões automatizadas

2.4. Segurança da informação e proteção de dados

O uso de IA frequentemente envolve o tratamento de dados, o que exige cuidados compatíveis com as políticas institucionais e com a legislação vigente.

De modo geral, espera-se que o servidor:

- avalie previamente a natureza das informações antes de utilizar ferramentas de IA;
- considere os riscos associados ao ambiente de uso (especialmente ferramentas externas);
- atue em conformidade com a LGPD e com as normas internas de segurança da informação.

Os sistemas de IA devem sempre serem compatíveis com a proteção de dados e a preservação do interesse público.

2.5. Adequação ao interesse público

A adoção da Inteligência Artificial deve estar alinhada às finalidades institucionais e aos princípios da Administração Pública.

Isso implica que:

- o uso da tecnologia deve gerar benefícios para a atividade institucional;
- as soluções adotadas devem respeitar critérios de imparcialidade, equidade e justiça.

A aplicação da tecnologia é de natureza instrumental, sendo validada pela sua capacidade de agregar eficácia aos resultados e aos processos institucionais.

2.6. Prudência e uso proporcional da tecnologia

Considerando que a Inteligência Artificial está em constante evolução, é necessário adotar uma postura cautelosa quanto à sua aplicação.

Nesse sentido:

- nem toda tarefa se beneficia do uso de IA;
- situações sensíveis ou complexas podem demandar maior interação humano ou até a não utilização da tecnologia;
- o uso deve ser proporcional ao risco envolvido e ao impacto potencial dos resultados.

A prudência evita que ganhos pontuais de eficiência resultem em riscos institucionais mais amplos.

3. Riscos Envolvidos na Adoção, Desenvolvimento e Uso de IA

O uso da Inteligência Artificial no ambiente institucional traz ganhos relevantes de eficiência e capacidade analítica, mas também introduz riscos que nem sempre são evidentes à primeira vista. Diferentemente de sistemas tradicionais, as ferramentas de IA operam com base em probabilidades e padrões, o que faz com que seus resultados possam variar em qualidade, precisão e adequação ao contexto.

Nesse cenário, a análise de riscos não deve ser compreendida como uma etapa isolada ou restritiva, mas como parte integrante do próprio uso da tecnologia. Em outras palavras, utilizar IA de forma adequada implica, necessariamente, reconhecer seus limites e agir para tratá-los no momento do uso.

Para facilitar essa compreensão, os principais riscos associados à IA são apresentados a seguir juntamente com a lógica de tratamento que deve orientar a atuação do servidor.

3.1. Confiabilidade das informações e validação do conteúdo

Um dos riscos mais característicos da IA é a capacidade de gerar alucinações, isto é:

- geração de informações factualmente incorretas ou desatualizadas;
- criação de referências inexistentes ou imprecisas;
- omissão de condições relevantes para a correta interpretação de normas;
- simplificação excessiva de temas complexos.

Esse comportamento pode induzir o usuário a confiar excessivamente no conteúdo apresentado, especialmente quando ele é bem estruturado.

Diante disso, a principal forma de tratamento é a validação crítica do conteúdo, que deve ir além da revisão textual. O servidor deve confrontar as informações com fontes confiáveis, verificar a aderência à legislação aplicável e ajustar o conteúdo ao contexto institucional.

O uso da IA, nesse caso, só se torna seguro quando acompanhado de validação.

3.2. Proteção de dados e controle da informação

Outro risco relevante está relacionado ao tratamento de dados, especialmente quando há interação com ferramentas externas à infraestrutura institucional.

Esse cenário pode envolver:

- exposição de dados pessoais ou institucionais em ambientes não controlados;
- compartilhamento indireto de informações sensíveis por meio de exemplos ou descrições;
- perda de controle sobre o armazenamento ou uso posterior dos dados;
- dificuldade de rastrear vazamentos ou usos indevidos.

A mitigação desse risco passa, antes de tudo, pela consciência sobre o ambiente de uso. O servidor deve avaliar previamente o tipo de informação envolvida e avaliar os dados submetidos conforme o nível de sensibilidade.

Além disso, deve-se adotar uma postura preventiva: quando houver dúvida sobre a adequação do uso, a alternativa mais segura é evitar a inserção de informações sensíveis.

3.3. Adequação ao contexto institucional

Os sistemas de IA podem não possuir conhecimento real do funcionamento ou fluxo interno da Autarquia, nem das especificidades normativas que orientam sua atuação. Por isso, determinadas respostas podem ser inadequadas ao contexto institucional, mesmo apresentando um formato correto.

Esse risco se manifesta na:

- aplicação de soluções genéricas a situações que exigem interpretação normativa específica;
- inadequação às competências e fluxos internos;
- desalinhamento com entendimentos consolidados ou padrões institucionais;

O tratamento desse risco depende da avaliação do servidor, que deve adequar o conteúdo gerado à realidade organizacional.

A IA pode sugerir caminhos, mas cabe ao servidor garantir que o resultado final esteja plenamente alinhado à atuação institucional.

3.4. Dependência tecnológica

O uso contínuo de uma única plataforma de Inteligência Artificial tende a produzir, ao longo do tempo, uma forma de “*vendor lock-in*”, reduzindo a viabilidade de migração para outros fornecedores ou soluções.

Esse aprisionamento se consolida de forma progressiva e pouco visível:

- fluxos de trabalho, prompts e rotinas são otimizados para o comportamento particular daquele modelo, tornando-os incompatíveis ou ineficientes em outros sistemas;
- o vocabulário, os padrões de resposta e os atalhos de interface passam a ser moldados pela solução de um único fornecedor;
- dados, históricos e contextos acumulados dentro da plataforma tornam-se barreiras de saída.

A mitigação passa pela adoção de camadas de abstração que desacoplam os fluxos de trabalho do fornecedor subjacente, preservando a portabilidade operacional independentemente de qual plataforma esteja sendo utilizada.

3.5. Transparência, rastreabilidade e capacidade de explicação

A utilização de IA pode introduzir dificuldades na compreensão de como determinado resultado foi produzido, especialmente quando seu uso não é devidamente registrado ou refletido.

Isso pode resultar em:

- dificuldade de justificar tecnicamente uma decisão;
- ausência de clareza sobre o processo de elaboração de documentos;
- fragilidade em situações de auditoria ou controle;
- percepção de falta de transparência institucional.

A forma de tratamento está na capacidade de explicação do uso realizado. O servidor deve ser capaz de descrever, em termos gerais, como utilizou a ferramenta e quais critérios orientaram a validação do resultado.

Sempre que a IA contribuir para atividades relevantes, é recomendável preservar elementos que garantam rastreabilidade, ainda que de forma simplificada.

3.6. Uso desproporcional ou inadequado da tecnologia

Nem toda tarefa se beneficia do uso de Inteligência Artificial, e seu emprego indiscriminado pode gerar mais riscos do que benefícios.

Esse tipo de situação ocorre quando:

- a IA é utilizada por conveniência, sem avaliação prévia de adequação;
- tarefas sensíveis são tratadas com ferramentas inadequadas;
- o nível de automação supera a complexidade da atividade;
- o ganho de eficiência não compensa o risco associado.

A mitigação desse risco passa pela avaliação prévia e pelo uso proporcional da tecnologia. O servidor deve considerar o impacto da atividade e ajustar o uso da IA conforme o seu nível de criticidade.

Quanto maior o risco envolvido, maior deve ser o controle humano sobre o processo.

De forma integrada, esses elementos demonstram que o uso seguro da Inteligência Artificial não depende apenas da tecnologia, mas também da forma como ela é utilizada. O tratamento de riscos não ocorre em um momento isolado, mas ao longo de toda a interação com a ferramenta.

3.7. Modelo de Avaliação de Riscos ao Longo do Ciclo de Uso da IA

Para complementar a abordagem conceitual de riscos apresentada anteriormente, adota-se um modelo estruturado de análise que considera as diferentes fases de utilização de sistemas de Inteligência Artificial sob a perspectiva do usuário.

Esse modelo permite organizar a identificação e o tratamento de riscos de forma mais sistemática, associando-os aos momentos concretos em que surgem durante o uso da tecnologia.

A jornada do usuário na utilização de soluções de IA pode ser compreendida em cinco fases principais, cada uma associada a riscos específicos que demandam atenção:

3.7.1. Extração e consolidação dos dados de entrada

Etapa em que o usuário realiza a seleção, coleta e organização dos dados que serão utilizados como insumo.

Nessa fase, destacam-se riscos relacionados à utilização de dados com vieses ou distorções, que não representem adequadamente a realidade, bem como ao uso indevido de informações confidenciais ou restritas sem autorização adequada.

3.7.2. Entrada ou exposição dos dados de entrada

Corresponde ao momento em que os dados são apresentados à ferramenta de IA.

Os principais riscos envolvem a formulação inadequada de instruções (prompts), uso de informações imprecisas ou incompletas e a utilização de ambientes inseguros, o que pode comprometer tanto a qualidade quanto a segurança dos resultados.

3.7.3. Processamento ou execução

Fase em que os dados são processados e analisados pela solução de IA.

Nessa etapa, surgem riscos associados ao uso de modelos com baixa transparência ou limitada capacidade de auditabilidade, bem como ao possível uso indevido ou não autorizado de dados por sistemas ou fornecedores.

3.7.4. Compilação e organização dos dados de saída

Etapa em que os resultados são apresentados e estruturados para o usuário.

Os riscos incluem a geração de conteúdos incompletos, incorretos ou de baixa qualidade, além do armazenamento ou compartilhamento dessas informações em ambientes inadequados ou inseguros.

3.7.5. Utilização dos dados e informações de saída

Momento em que os resultados são efetivamente utilizados em atividades institucionais.

O principal risco consiste na utilização desses dados sem validação adequada, especialmente em decisões ou ações que possam impactar direitos, serviços ou processos relevantes.

A análise integrada dessas fases reforça que o risco no uso da IA não está concentrado apenas no resultado final, mas distribuído ao longo de toda a interação com a tecnologia. Isso exige uma atuação contínua, crítica e proporcional por parte do servidor em cada etapa do processo.

3.8. Modelo de Avaliação de Riscos associado ao desenvolvimento de projetos

Nos casos de adoção e desenvolvimento de produtos de IA, recomenda-se utilizar o framework de [Autoavaliação de Impacto Ético da Inteligência Artificial no Setor Público \(AIE\)](#), elaborado pelo Ministério da Gestão e Inovação no Serviço Público (MGI).

4. Considerações Finais

A adoção da Inteligência Artificial na Administração Pública representa um avanço relevante na forma como o trabalho é realizado, permitindo maior eficiência, organização e capacidade analítica.

No entanto, seu uso exige a consolidação de uma cultura institucional baseada na responsabilidade, na consciência de riscos e no compromisso com o interesse público.

Este documento não tem como objetivo restringir o uso da tecnologia, mas orientar sua aplicação de forma segura, consistente e alinhada aos princípios administrativos. Ele deve ser lido em conjunto com o Guia de Boas Práticas, formando um conjunto integrado de referências para o servidor.

Mais do que conhecer diretrizes, espera-se que o usuário da ferramenta desenvolva uma postura crítica e reflexiva, capaz de adaptar o uso da IA às diferentes situações do cotidiano, sempre com responsabilidade sobre o resultado final.

A efetividade das diretrizes depende da forma como são incorporadas à prática. Nesse sentido, o servidor não é apenas usuário da tecnologia, mas protagonista da sua utilização responsável no âmbito da Administração Pública.

5. Referências

BRASIL. Ministério da Gestão e da Inovação em Serviços Públicos. Secretaria de Governo Digital. **Framework para a Autoavaliação de Impacto Ético em Inteligência Artificial no Setor Público Federal (AIE)**. Brasília: MGI, 2024. Disponível em: <https://www.gov.br/governodigital/pt-br/infraestrutura-nacional-de-dados/inteligencia-artificial-1/publicacoes/framework-para-a-autoavaliacao-de-impacto-etico-em-inteligencia-artificial-no-setor-publico-federal>. Acesso em: maio 2026.

BRASIL. Ministério da Gestão e da Inovação em Serviços Públicos. Secretaria de Governo Digital. **Inteligência Artificial**. Brasília: MGI, [2026?]. Disponível em: <https://www.gov.br/governodigital/pt-br/infraestrutura-nacional-de-dados/inteligencia-artificial-1>. Acesso em: maio 2026.

UNESCO. **Recomendação sobre a Ética da Inteligência Artificial**. Brasília: UNESCO, 2022. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000381137_por. Acesso em: maio 2026.