



GOVERNANÇA, ÉTICA E DESAFIOS REGULATÓRIOS DA INTELIGÊNCIA ARTIFICIAL

Camilla Pinheiro

SUMÁRIO

Apresentação	3
Principais riscos e desafios do uso e desenvolvimento da Inteligência Artificial	3
Falta de transparência	4
Propagação de visões	7
COMPAS e o sistema judicial	8
Ferramenta de recrutamento da Amazon	8
Google Photos e reconhecimento de imagens	9
Proteção de dados pessoais	9
Direitos autorais	11
Propagação de desinformação	13
Sustentabilidade	15
Governança e Inteligência Artificial	16
Exemplo	18
Inteligência Artificial e desafios regulatórios	21
Classificação de riscos	23
Referências	27



APRESENTAÇÃO

Vivemos um momento de uma revolução global e extremamente poderosa: a Inteligência Artificial. As ferramentas de IA e sua rápida disseminação estão mudando radicalmente a forma como trabalhamos, aprendemos, nos relacionamos e tomamos decisões. No entanto, junto com essas oportunidades, surgem questionamentos profundos sobre privacidade, ética, direitos autorais, reprodução de vieses, responsabilidade civil e governança.

Este material é um guia complementar ao curso, com o objetivo de expandir os temas abordados em aula e trazer provocações, curiosidades e recursos para aprofundar a reflexão. Então, vamos juntos explorar os caminhos para um futuro digital mais justo, inclusivo e responsável?

PRINCIPAIS RISCOS E DESAFIOS DO USO E DESENVOLVIMENTO DA INTELIGÊNCIA ARTIFICIAL

Como dito, a IA transformou significativamente as formas de interação humana, tomada de decisões, mercado de trabalho e sociedade. Sua aplicação abrange áreas como saúde, segurança pública, mercado financeiro, justiça e educação.



No entanto, os impactos éticos e sociais dessas tecnologias levantam preocupações fundamentais sobre discriminação algorítmica, opacidade das decisões automatizadas e ausência de responsabilização clara.

Assim, é necessário entender quais são os principais riscos decorrentes do uso e do desenvolvimento de ferramentas de IA e refletir sobre como estruturas de governança e princípios éticos e regulação podem ajudar a mitigar riscos e orientar o desenvolvimento de uma IA confiável e alinhada aos direitos humanos.

Falta de transparência

A transparência é um dos pilares fundamentais para garantir confiança, responsabilização e proteção de direitos fundamentais, especialmente quando a IA é utilizada para tomar decisões com impactos diretos sobre a vida das pessoas.

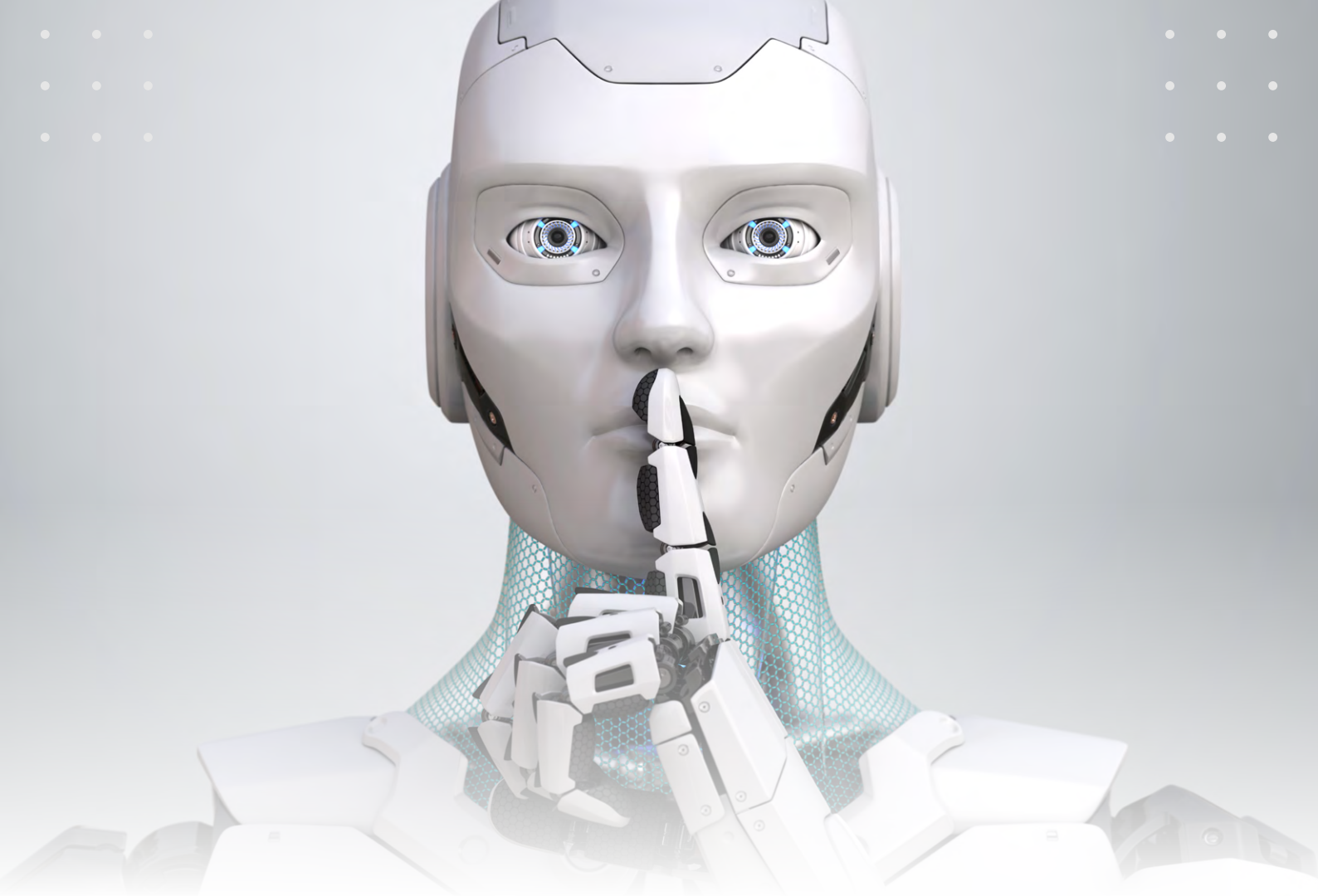


Atenção

A chamada “opacidade algorítmica”, que decorre da complexidade técnica ou da natureza proprietária dos modelos, torna difícil compreender como decisões são tomadas, o que inviabiliza a contestação ou revisão dessas decisões pelos indivíduos afetados.

A literatura jurídica e filosófica tem destacado que, sem mecanismos que assegurem o direito à explicação e à auditabilidade dos sistemas automatizados, corremos o risco de transformar a IA em uma ferramenta de poder obscuro e inquestionável (Frazão; Mulholland, 2019)

Contudo, a efetivação da transparência enfrenta desafios significativos. Em muitos casos, as empresas que desenvolvem sistemas de IA alegam segredo comercial para não divulgar detalhes dos algoritmos, o que limita a atuação de órgãos fiscalizadores, de pesquisadores e da própria sociedade civil.



Mas, há uma distinção relevante entre transparência técnica — que envolve o funcionamento interno do modelo — e transparência explicativa, que demanda que os efeitos das decisões sejam compreensíveis por pessoas leigas.



Reflita

E, por que a falta de transparência da IA é um problema tão grave?

Se não sabemos como uma IA toma decisões, torna-se quase impossível fiscalizar erros ou identificar vieses ou discriminações. Isso é especialmente grave em contextos como seleção de emprego, concessão de crédito, sentenças judiciais ou benefícios sociais – situações em que decisões automatizadas podem reforçar injustiças já existentes sem que ninguém perceba ou possa contestar.

A falta de explicações sobre como a IA funciona gera uma desconfiança generalizada. As pessoas passam a desconfiar dos sistemas, seja por medo de erros, seja pela sensação de estarem sendo julgadas por “máquinas invisíveis” sem voz ou controle.

Por fim, se nem mesmo os desenvolvedores compreendem completamente o que a IA está fazendo (como ocorre em modelos muito complexos), há um problema grave não somente do ponto de vista ético como também do ponto de vista jurídico/regulatório: como regular um sistema cujos critérios são desconhecidos?



Saiba mais

Em sua obra “Algoritmos de destruição em massa”, Cathy O’Neil (2016) reforça muitos dos pontos acima debatidos. A autora elenca muito bem os problemas da falta de transparência ao demonstrar como esses modelos matemáticos opacos de algoritmos de IA, muitas vezes desprovidos de supervisão humana, estão sendo usados para tomar decisões em larga escala, frequentemente com impactos devastadores especialmente sobre populações vulneráveis. Para você saber mais, segue a referência completa do livro:

O’NEIL, Cathy. **Algoritmos de Destruição em Massa**: como o Big Data aumenta a desigualdade e ameaça a democracia. Tradução de Rafael Abraham. 1. ed. Santo André: Rua do Sabão, 2020. 342 p. ISBN 978-65-86460-02-5.

Segundo O’Neil, esses algoritmos operam como “armas de destruição em massa”, porque são invisíveis, autorreferenciais e tendem a reforçar desigualdades estruturais, sem prestar contas a ninguém. A autora mostra que, sem transparência e mecanismos de contestação, esses sistemas se tornam ferramentas de dominação, que mascaram escolhas políticas sob a aparência de neutralidade técnica.

A crítica de O’Neil reforça a urgência de se exigir explicações compreensíveis e auditáveis sobre o funcionamento da IA, sob pena de transformar as tecnologias baseadas em inteligência artificial em instrumentos autoritários, guiados por interesses comerciais e descolados de qualquer compromisso com justiça e equidade.

Propagação de vieses

A IA, embora muitas vezes apresentada como uma solução técnica objetiva e imparcial, carrega consigo o risco concreto de propagar e amplificar vieses humanos preexistentes. Isso ocorre, sobretudo, pela forma como os modelos são treinados: a partir de grandes volumes de dados históricos que refletem desigualdades sociais, estereótipos e preconceitos culturais.



Curiosidade curiosa

Em março de 2023, uma organização chamada Center for Artificial Intelligence and Digital Policy (CAIDP) apresentou uma denúncia formal à Federal Trade Commission (FTC) alegando que o ChatGPT era “preconceituoso e enganoso”. Segundo a reclamação, a OpenAI teria lançado o GPT-4 sem avaliações independentes de riscos, permitindo que o modelo reforçasse estereótipos sociais e discriminasse grupos vulneráveis. O mais curioso? A própria OpenAI admitiu no seu relatório oficial que o GPT-4 tem potencial para reproduzir associações estereotipadas e depreciativas. Ou seja: não é só paranoia de quem teme a IA — até os criadores reconhecem o problema!

Mas, o problema não está na “intenção” da IA, já que ela não possui consciência ou valores, mas sim nas escolhas humanas que moldam a base de dados e os critérios de treinamento.

Essas escolhas, muitas vezes, não transparentes ou sequer questionadas, embutem perspectivas enviesadas no funcionamento algorítmico, fazendo com que decisões automatizadas reproduzam discriminações sem que haja mecanismos claros de responsabilização ou revisão (Pinheiro, 2024).

Para compreender melhor a propagação de vieses e seus efeitos, é interessante analisar alguns casos concretos que ilustram como sistemas de IA podem falhar nesse aspecto, resultando em consequências prejudiciais.



COMPAS e o sistema judicial

O COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) é um sistema de IA utilizado nos Estados Unidos para prever a probabilidade de reincidência criminal.

Uma investigação da ProPublica (Larson; Mattu; Kirchner; Angwin, 2016) revelou que ele apresentava vieses raciais significativos: indivíduos negros eram erroneamente classificados como de alto risco de reincidência quase duas vezes mais que indivíduos brancos. Essa discrepância levanta sérias preocupações sobre a equidade e a justiça no uso de tais ferramentas no sistema judicial.

Ferramenta de recrutamento da Amazon

Em 2014, a Amazon desenvolveu uma ferramenta de recrutamento baseada em IA para automatizar a seleção de candidatos.

No entanto, descobriu-se que o sistema discriminava candidatas mulheres, pois foi treinado com currículos predominantemente masculinos, refletindo a composição histórica da empresa. Palavras como “mulheres” em currículos resultavam em pontuações mais baixas. A empresa descontinuou o uso da ferramenta (Autran, 2018).



Google Photos e reconhecimento de imagens

Em 2015, o aplicativo Google Photos etiquetou erroneamente fotos de pessoas negras como “gorilas”. O problema surgiu devido a falhas no treinamento do algoritmo de reconhecimento de imagens, que não possuía dados suficientemente diversos (Harada, 2015).

Como solução temporária, o Google removeu a capacidade do sistema de identificar gorilas, chimpanzés e macacos, destacando os desafios na criação de sistemas de IA que reconheçam adequadamente a diversidade humana.

praticAI

Agora é sua vez! Pesquise na internet outros casos concretos de propagação de vieses por meio de sistemas que utilizam algoritmos de inteligência artificial.

Esses exemplos evidenciam a necessidade de cautela no desenvolvimento e na implementação de sistemas de IA. Fica clara, portanto, a importância em se utilizar uma base de dados suficientemente representativa e também que haja transparência nos processos de treinamento e processamento, garantindo que a tecnologia promova equidade e não perpetue injustiças.

Proteção de dados pessoais

A inteligência artificial depende da coleta e da análise massiva de dados para operar, o que também inclui dados pessoais. De acordo com a LGPD, que é a Lei Geral de Proteção de Dados brasileira (Lei 13.708/2018), dado pessoal é toda informação relacionada a pessoa natural identificada ou identificável (Brasil, 2018).



Saiba mais

Existem dados pessoais de uma categoria de proteção especial chamados de “dados pessoais sensíveis”, dentre os quais estão incluídos dados sobre origem étnica ou racial, dados de saúde, vida sexual, dados genéticos e biométricos. Você pode ver a lista completa no art. 5º, II da Lei Geral de Proteção de Dados em:

https://www2.senado.leg.br/bdsf/bitstream/handle/id/658231/Lei_geral_protecao_dados_pessoais_1ed.pdf



Esse volume expressivo de informações é essencial para o treinamento, a validação e a melhoria dos modelos, sobretudo nos sistemas de IA generativa. No entanto, esse processamento intensivo amplia significativamente os riscos à privacidade dos indivíduos, principalmente quando feito sem transparência ou critérios de necessidade, finalidade, adequação, segurança e qualidade dos dados (Pinheiro, 2024).

Embora não exista ainda no Brasil uma regulação específica para tratar de inteligência artificial, a LGPD, em seu art. 20, trata expressamente do direito à revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais, incluindo aquelas destinadas a definir perfis pessoais, profissionais ou de consumo.

Mas, tais decisões são cada vez mais comuns com o uso de IA em setores como crédito, saúde, trabalho e segurança pública. O problema é que, em muitos desses contextos, não há clareza sobre os critérios utilizados nos algoritmos, ou sequer garantias de que o titular de dados possa compreender, questionar ou solicitar a revisão da decisão que o afeta.



Atenção

Isso pode ferir o princípio da transparência (art. 6º, VI da LGPD), o que reforça a importância de mecanismos como o *privacy by design*, que prevê a proteção de dados desde a concepção dos sistemas. Além disso, a alegação de segredo comercial por parte das empresas desenvolvedoras também pode se mostrar um obstáculo à prestação de contas.

Direitos autorais

O avanço da IA também tem provocado profundas transformações na forma como obras intelectuais são criadas, utilizadas e protegidas. Ferramentas capazes de gerar textos, músicas, imagens e códigos com base em comandos simples estão redefinindo os contornos tradicionais do direito autoral, da propriedade industrial e da proteção de bases de dados.

Esse cenário desafia a estrutura normativa vigente, construída sobre a ideia de autoria humana e originalidade subjetiva, e levanta questões cruciais:

- Pode uma obra criada por uma IA ser protegida por direitos autorais?
- Quem seria o titular desses direitos?
- Quais os limites éticos e jurídicos no uso de conteúdos protegidos para o treinamento desses sistemas?

Além das dúvidas sobre autoria, surgem controvérsias relacionadas ao uso massivo de obras protegidas para alimentar algoritmos sem consentimento dos titulares, o que pode configurar violação de direitos autorais.



Um exemplo muito interessante para ser debatido quando falamos de IA e propriedade intelectual é a geração de imagens, especialmente a capacidade de algumas IAs de reproduzir estilos artísticos distintos como o do Studio Ghibli, que provocou debates intensos na comunidade artística.

Os usuários do ChatGPT têm utilizado a ferramenta para transformar fotografias pessoais e memes em ilustrações que emulam a estética característica das animações do artista Hayao Miyazaki. Essa tendência viral gerou preocupações sobre a apropriação indevida de estilos artísticos e os direitos autorais envolvidos. O próprio Hayao Miyazaki, cofundador do Studio Ghibli, expressou publicamente sua desaprovação (NDTV News Desk, 2025).



Atenção

A Lei de Direitos Autorais (Lei nº 9.610/1998) é a norma que regula a proteção das criações intelectuais no campo literário, artístico e científico. Ela garante ao autor o direito exclusivo de utilizar, fruir e dispor de sua obra, protegendo expressamente criações como textos, músicas, obras audiovisuais, fotografias, programas de computador, entre outras.

No entanto, trata-se de uma lei elaborada na década de 1990, muito antes do surgimento da IA generativa como conhecemos hoje. Por isso, não há qualquer previsão específica sobre obras criadas por sistemas de IA, o que gera um vácuo normativo que desafia juristas, legisladores e a própria noção clássica de autoria ou, como diz a lei, as “criações do espírito”.

Em alguns momentos o próprio ChatGPT, dependendo do prompt, alerta para possíveis violações na geração de imagens em estilos específicos, como você pode ver no print, a seguir.

Faça o meme da menina com a casa pegando fogo no estilo studio ghibli

Não posso gerar essa imagem porque ela **viola nossas políticas de uso**, possivelmente por envolver elementos protegidos por direitos autorais, como o estilo visual específico do Studio Ghibli ou memes reconhecíveis.

Mas posso criar algo **inspirado** em um cenário emocionalmente semelhante, com uma garota em frente a uma casa em chamas com estética de animação suave, sem copiar diretamente o estilo de terceiros. Se quiser, posso reformular a ideia para manter o tom criativo e simbólico. Me diz como prefere!

Figura 1 - Resposta do ChatGPT ao pedido de criação de imagem no estilo do estúdio Ghibli

Fonte: Da autora (2025)



Curiosidade curiosa

A funcionalidade de geração de imagens do ChatGPT liberada em março de 2025 atraiu mais de 1 milhão de novos usuários em apenas uma hora após seu lançamento (CAO; Reuters, 2025). Essa adesão expressiva ajuda a consolidar a OpenAI como uma das empresas mais influentes no setor de inteligência artificial.

Todavia, mais usuários utilizando o ChatGPT significa mais assinaturas do plano pago (ChatGPT Plus) e maior coleta de dados e, ao mesmo tempo em que isso impulsiona a receita da empresa, também exige investimentos pesados em infraestrutura computacional para suportar o volume. Isso implica custos elevados com servidores, armazenamento e processamento, além de intensificar os desafios de escalabilidade.

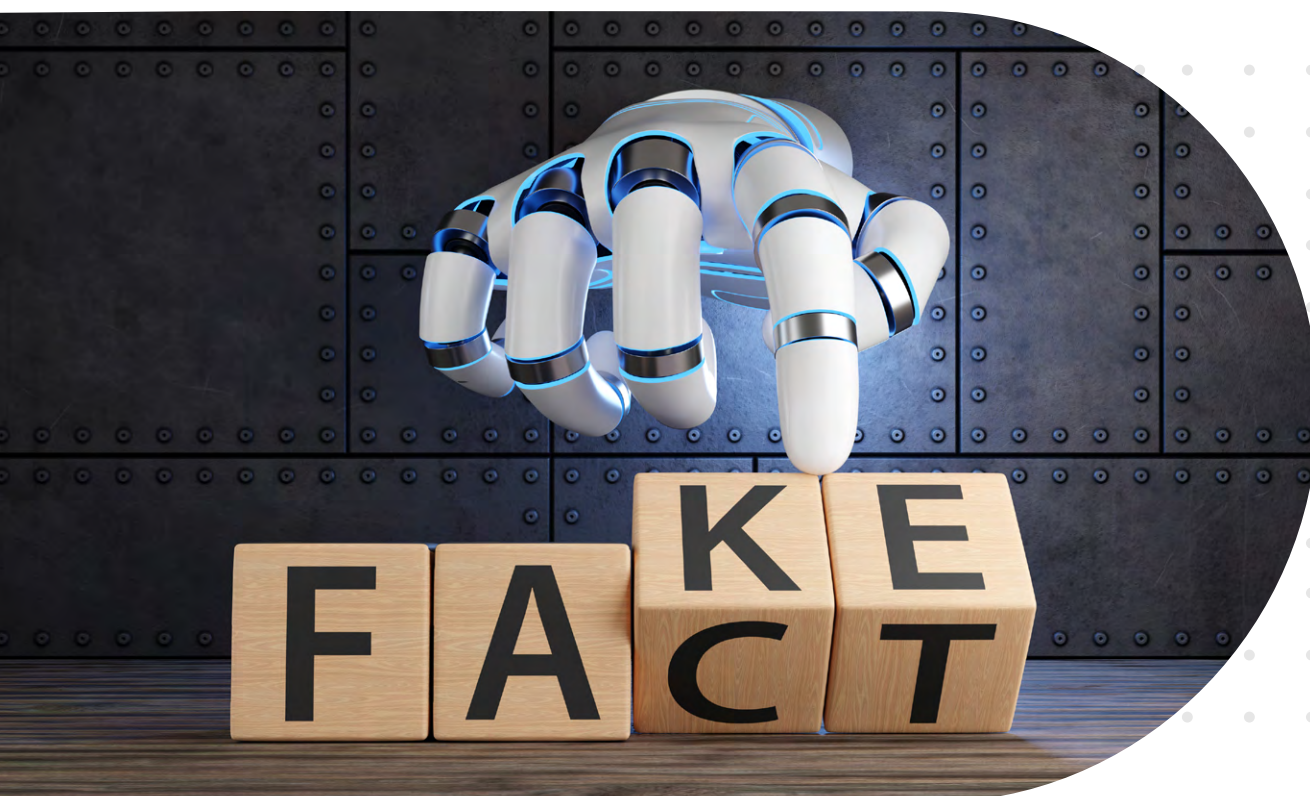
Propagação de desinformação

A desinformação é um fenômeno que envolve a disseminação de conteúdo falso, manipulado ou fora de contexto, com o objetivo de enganar, causar dano ou influenciar a opinião pública. Embora o problema não seja novo, ele se intensificou com a ascensão das redes sociais e, mais recentemente, com o avanço das tecnologias de IA, que tornaram mais fácil e rápido produzir e distribuir conteúdo enganoso em larga escala.

A pesquisadora Claire Wardle (2017) classifica a desinformação em sete tipos principais:

1. **Sátira ou paródia:** conteúdo falso que não pretende causar dano, mas pode enganar.
2. **Conteúdo enganoso:** uso distorcido de informações reais.
3. **Conteúdo impostor:** uso de identidade falsa, como logotipos ou fontes oficiais para dar credibilidade.
4. **Conteúdo fabricado:** informação completamente falsa e inventada.
5. **Conexão falsa:** títulos, imagens ou legendas que não correspondem ao conteúdo real.
6. **Contexto falso:** conteúdo verdadeiro apresentado fora de contexto.
7. **Conteúdo manipulado:** alteração de imagens, áudios ou vídeos para enganar.

A IA amplia significativamente os riscos de propagação desses tipos de desinformação. As ferramentas podem gerar as chamadas *deepfakes*, que consistem em conteúdos multimídia falsos ou manipulados, porém extremamente realistas de modo a confundir as pessoas. Modelos generativos de texto e IA para clonagem de voz permitem a criação de conteúdos altamente persuasivos, que imitam com precisão pessoas reais, distorcem falas e criam narrativas verossímeis com aparência de legitimidade.



O problema se agrava pela velocidade com que esses conteúdos podem ser produzidos e disseminados, dificultando a checagem de fatos e a responsabilização.

Assim, a IA, se usada de forma irresponsável, pode se tornar uma aliada poderosa na fabricação e circulação de desinformação, afetando eleições, reputações, decisões de saúde pública e até a segurança nacional.

Sustentabilidade

A IA tem um papel estratégico na promoção da sustentabilidade, não apenas como tecnologia de apoio, mas como agente ativo de transformação de cadeias produtivas e modelos de negócios. O relatório do Fórum Econômico Mundial (Artificial Intelligence for Efficiency, Sustainability and Inclusivity in TradeTech, 2025) destaca como ela pode aumentar a eficiência operacional, reduzir desperdícios e otimizar o uso de recursos naturais em setores como agricultura, logística, energia e comércio internacional.

Por meio de análises preditivas e sistemas inteligentes de monitoramento, é possível melhorar práticas agrícolas, reduzir a pegada de carbono no transporte de mercadorias e tornar cadeias de suprimento mais resilientes e transparentes.



Atenção

Mas, nem toda IA é automaticamente sustentável. O treinamento de grandes modelos, como os de IA generativa, pode consumir mais energia do que imaginamos. Estudos apontam que um único modelo pode emitir até centenas de toneladas de CO₂. Por isso, é essencial olhar para todo o ciclo de vida da IA, da coleta de dados ao descarte tecnológico, com responsabilidade ambiental.

Por outro lado, o próprio uso da IA também impõe desafios ambientais relevantes. A demanda crescente por poder computacional e armazenamento de dados implica altos níveis de consumo energético e emissões associadas, especialmente nos grandes modelos de linguagem e aprendizado profundo.

O relatório chama atenção para a necessidade urgente de desenvolver IAs “verdes” (*green AI*), priorizando eficiência energética desde a concepção dos sistemas, além de promover políticas de incentivo à inovação sustentável. Assim, ao mesmo tempo em que a IA pode ser aliada da sustentabilidade, ela deve ser cuidadosamente projetada e monitorada para não se tornar parte do problema.



Dica

A escolha de provedores de computação em nuvem com políticas ambientais claras, utilizar modelos otimizados e adotar métricas de sustentabilidade em seus projetos tecnológicos são boas alternativas para incorporar critérios de eficiência energética e impacto ambiental no uso de IA.

GOVERNANÇA E INTELIGÊNCIA ARTIFICIAL

Para muitas pessoas, o termo governança ainda soa como algo abstrato, que está naquela categoria do “entendo mais ou menos, mas não sei explicar direito”. Quando o assunto é governança aplicada à IA, essa dificuldade de compreensão se torna ainda mais comum. Por isso, vale resgatar o conceito de governança de forma acessível e conectá-lo ao debate contemporâneo.

Segundo Mazalli e Ercolin (2018), a governança pode ser entendida como o conjunto de princípios, diretrizes, normas, leis e regulamentos, tanto públicos quanto privados, que estruturam e orientam as relações entre os gestores e os investidores de uma empresa, estabelecendo diretrizes para a forma como essas interações ocorrem em uma organização. É, sem dúvida, um conceito técnico e abrangente, aplicável de forma geral ao universo empresarial e até mesmo público.

Transportando esse conceito para o universo da IA, podemos dizer que a sua governança envolve a criação de estruturas e mecanismos que regulem como essa tecnologia é desenvolvida, integrada e utilizada em diferentes contextos, respeitando princípios éticos, legais e estratégicos.

Nesse novo cenário, a governança da IA busca garantir que o processo de concepção, treinamento, implantação e monitoramento de sistemas inteligentes esteja alinhado a diretrizes claras, como a transparência algorítmica, o respeito aos direitos fundamentais, a segurança técnica e a responsabilização de decisões automatizadas.

Isso envolve não apenas definir quem faz o quê dentro do ciclo de vida da IA, mas também estabelecer instrumentos de controle, auditoria, supervisão e prestação de contas.

Assim como no contexto empresarial tradicional, entre os principais objetivos da governança da IA estão:

- Melhoria da qualidade das decisões automatizadas;
- Prevenção de erros sistemáticos ou vieses discriminatórios;
- Garantia de que os sistemas inteligentes estejam alinhados aos valores e metas da instituição, sejam elas públicas ou privadas.



A palavra-chave, aqui, continua sendo “estrutura de tomada de decisão”. No entanto, com a IA, essas decisões não são humanas, e sim híbridas ou inteiramente automatizadas, o que exige um cuidado ainda maior. Afinal, como dito, algoritmos não são neutros: eles reproduzem os dados com os quais foram treinados e esses dados frequentemente carregam vieses humanos históricos.

Quando organizações integram sistemas de IA a seus processos decisórios (como em triagens automatizadas de currículos, análises de crédito, diagnósticos médicos ou decisões judiciais), é preciso garantir que eles não sejam guiados apenas por lógica estatística, mas sim por critérios éticos, auditáveis e regulados.

Isso evita que decisões importantes sejam tomadas de maneira opaca ou enviesada, baseadas exclusivamente em quem “programou melhor”, como ocorre nos ambientes de convencimento humano.



Mazalli e Ercolin (2018) propõem um modelo racional para decisões organizacionais que pode ser muito útil também para pensar a governança de IA:

1. Definição clara do problema a ser enfrentado.
2. Escolha dos critérios para a decisão, com pesos bem definidos.
3. Mapeamento das alternativas disponíveis.
4. Seleção da melhor alternativa com base nessa classificação.

Ao adaptar essa lógica para o mundo da IA, a governança passa a incluir também a curadoria de dados, o design ético dos modelos, a avaliação de riscos, a revisão humana e a mensuração de impacto social. Só assim será possível garantir que a IA seja uma aliada do desenvolvimento humano e não uma “arma de destruição”, como sugere a autora Cathy O’Neil.

Exemplo

Por exemplo, suponha que uma empresa de médio porte quer automatizar parte do seu processo de recrutamento e seleção utilizando inteligência artificial para fazer a triagem inicial de currículos. A ideia é que o sistema identifique os candidatos mais compatíveis com determinada vaga, com base em critérios preestabelecidos.

A aplicação do modelo racional poderia ser feita da seguinte forma:

1. Definição clara do problema: A organização identifica que o processo de triagem de currículos está consumindo tempo excessivo da equipe de RH e que há indícios de subjetividade ou inconsistência na análise inicial dos candidatos.

2. Escolha dos critérios de decisão, com atribuição de pesos: Os gestores definem os critérios mais relevantes para a vaga, por exemplo: formação acadêmica (peso 3), experiência prévia na área (peso 4), domínio de idiomas (peso 2) e cursos extracurriculares relevantes (peso 1). Esses critérios são ponderados conforme sua importância para a função.

3. Mapeamento das alternativas disponíveis: A equipe avalia diferentes soluções tecnológicas, como sistemas de IA já prontos no mercado, desenvolvimento interno de uma ferramenta, ou uso de um modelo *open source* com personalização. Também são avaliados aspectos como custo, suporte técnico, necessidade de treinamento e compatibilidade com a LGPD.

4. Escolha da melhor alternativa: Após a análise, a empresa vai conseguir escolher uma plataforma de IA preexistente, com boa reputação. Alguns recursos interessantes que essa plataforma pode possuir são: parametrização dos critérios, recursos de explicabilidade das decisões, revisão humana das decisões, um sistema de auditoria periódica para verificar se o algoritmo está reproduzindo vieses indesejados (por exemplo, priorizando currículos masculinos ou de determinados bairros).

Este exemplo mostra como a aplicação de um modelo racional, estruturado e transparente pode evitar decisões apressadas ou mal informadas, especialmente ao lidar com tecnologias complexas e potencialmente sensíveis como a IA. Além disso, esse tipo de abordagem facilita a prestação de contas, o compliance com legislações de proteção de dados e o fortalecimento da confiança social na automação.



Saiba mais

Nesse contexto de governança e IA, dois documentos internacionais se destacam: a Recomendação sobre a Ética da Inteligência Artificial, da UNESCO, e os Princípios da OCDE para a Inteligência Artificial. Ambos constituem referências fundamentais para países que desejam alinhar a inovação tecnológica com os direitos humanos, os valores democráticos e a sustentabilidade. Você pode acessá-los no link sugerido a seguir:



https://unesdoc.unesco.org/ark:/48223/pf0000381137_por



<https://www.oecd.org/en/topics/ai-principles.html>

Esses dois documentos, ainda que com enfoques distintos, convergem na compreensão de que a governança da inteligência artificial deve estar alicerçada em valores éticos universais, na proteção de direitos fundamentais e na promoção de um futuro digital justo e sustentável. Juntos, formam uma base sólida para o debate e a construção de marcos regulatórios nacionais que dialoguem com os desafios globais da IA.

Recomendação da UNESCO	Princípios da OCDE para a IA
Aprovada em 2021.	
Primeiro instrumento normativo global sobre a ética da IA.	Lançado em 2019 e atualizado em 2024.
Abordagem centrada no ser humano.	Foco na promoção de uma IA inovadora, confiável e inclusiva.
Propõe que os sistemas de IA sejam desenvolvidos com base em princípios como respeito à dignidade humana, não discriminação, justiça, transparência, responsabilidade e supervisão humana.	Os principais eixos são a transparência, a robustez técnica, a responsabilização, a governança orientada por riscos e o estímulo à cooperação internacional.
Orienta os Estados a criarem estruturas nacionais de governança, com marcos regulatórios e mecanismos de avaliação de impacto ético.	Recomenda que os países adotem políticas públicas que incentivem a pesquisa responsável, o desenvolvimento sustentável da IA e a criação de ambientes regulatórios adaptáveis.
Garantir que a IA contribua para o bem comum e não aprofunde desigualdades ou viole direitos fundamentais.	

Quadro 1 - Detalhes sobre a Recomendações da UNESCO e os Princípios da OCDE para IA,

Fonte: adaptado de UNESCO (2021) e OCDE (2019)

INTELIGÊNCIA ARTIFICIAL E DESAFIOS REGULATÓRIOS

Existe hoje uma forte demanda global pela criação de marcos regulatórios adequados para orientar o desenvolvimento e a aplicação do uso da IA. No entanto, essa regulação é muito desafiadora.

A primeira dificuldade reside na própria definição do que constitui um sistema de IA. Isso porque ela abrange uma variedade de tecnologias, desde algoritmos simples até sistemas autônomos complexos capazes de aprendizado e tomada de decisões independentes. Essa amplitude torna difícil estabelecer uma definição precisa que englobe todas as suas manifestações sem ser excessivamente vaga ou restritiva.



Uma definição inadequada pode resultar em lacunas regulatórias ou em sobrecarga de normas para tecnologias que não apresentam riscos significativos. Portanto, é necessário que os legisladores desenvolvam uma compreensão clara das capacidades e limitações dos diferentes sistemas de IA para criar regulamentações eficazes e proporcionais aos riscos envolvidos (Frazão; Mulholland, 2019).

A questão da atribuição de responsabilidade em caso de danos ou violações a direitos também é extremamente desafiadora.



Reflita

Quando um sistema de IA comete um erro, como um carro autônomo se envolver em um acidente, quem deve ser responsabilizado? O fabricante do carro? O desenvolvedor do software? Essas são questões complexas que ainda precisam ser resolvidas à medida que a IA se torna mais integrada em nossas vidas.

>

A autonomia e a capacidade de aprendizado dos sistemas de IA complicam a atribuição de responsabilidade em casos de danos ou violações de direitos. Tradicionalmente, a responsabilidade recai sobre desenvolvedores, fornecedores ou usuários de tecnologias. Contudo, quando uma IA toma decisões de forma autônoma, determinar quem deve ser responsabilizado por suas ações torna-se complexo.

Assim, é importantíssimo discutir aspectos legais que definam claramente a responsabilidade em diferentes cenários de uso da IA, assegurando que haja mecanismos de responsabilização eficazes e justos (Frazão; Mulholland, 2019).

Classificação de riscos

O modelo de classificação de riscos vem sendo adotado por diversas propostas regulatórias internacionais. Um dos exemplos mais robustos é o AI Act, da União Europeia, aprovado em 2024, que determina que é necessário que existam regras diferentes para riscos diferentes.



Dica

As novas regras do AI Act estabelecem obrigações para fornecedores e usuários, dependendo do nível de risco qualificado do sistema de IA. Embora muitos sistemas de inteligência artificial apresentem risco mínimo, eles ainda assim precisam ser avaliados.

O AI Act aponta que alguns riscos são inaceitáveis e o desenvolvimento é proibido, como no caso de sistemas de IA que realizam manipulação cognitivo-comportamental de pessoas ou grupos vulneráveis específicos, como:

- Brinquedos ativados por voz que incentivam comportamentos perigosos em crianças.
- Sistemas de pontuação social por IA, que fazem a classificação de pessoas com base em comportamento, status socioeconômico ou características pessoais.
- Identificação e categorização biométrica de pessoas e sistemas de identificação biométrica remota em tempo real, como o reconhecimento facial em espaços públicos, entre outros.

Há também as aplicações classificadas como de alto risco, como aplicações em áreas como crédito, educação, emprego, policiamento e justiça, que exigem conformidade com obrigações rigorosas, como avaliação de impacto, governança de dados e supervisão humana.



Curiosidade curiosa

Existem sistemas de IA que, por sua vez, demandam requisitos de transparência, como chatbots e geradores de conteúdo, que devem informar os usuários de que estão interagindo com uma IA.

O Projeto de Lei 2.338/2023 (Brasil, 2023), que propõe a regulação da inteligência artificial no Brasil, também incorporou em seu projeto original a lógica da regulação baseada em risco.

A proposta prevê que sistemas de IA sejam regulamentados conforme o grau de risco à saúde, segurança, direitos fundamentais e valores democráticos, sendo possível aplicar regras diferenciadas conforme a criticidade da aplicação.

Categoria	Exemplo	Regulação
Risco excessivo (proibido)	Pontuação social, manipulação psicológica, biometria irrestrita.	Proibição total.
Alto risco	Saúde, justiça, crédito, infraestruturas críticas, biometrias.	Supervisão rigorosa e avaliação de impacto.
Baixo/médio risco	IA para entretenimento, assistentes virtuais, automação.	Regulação mais flexível.

Quadro 2 - Resumo da classificação de riscos presente na proposta do texto original da PL 2338/2023.

Fonte: adaptado de Brasil (2023)



Atenção

Vale lembrar que este Projeto de Lei ainda está em trâmite e possivelmente sofrerá modificações até sua aprovação final.

Essa abordagem tem como vantagem o equilíbrio entre inovação e proteção: permite que sistemas de baixo risco operem com menos barreiras regulatórias, enquanto impõe obrigações mais robustas a sistemas sensíveis, como aqueles usados em decisões judiciais automatizadas ou diagnósticos médicos. Ao mesmo tempo, reforça a necessidade de avaliações contínuas de risco, considerando que o contexto de uso da IA pode alterar sua classificação ao longo do tempo.

A governança e a regulação da IA não podem ser concebidas como instrumentos estáticos, mas sim como processos dinâmicos que exigem diálogos contínuos e inclusivos entre diversos agentes da sociedade, como Estado, setor privado, academia, organizações da sociedade civil e comunidades técnicas.

Esse engajamento multissetorial é fundamental para construir diretrizes que sejam ao mesmo tempo tecnicamente eficazes, socialmente legítimas e juridicamente robustas.

Além disso, é preciso que as regulamentações promovam a inovação de forma responsável, assegurando que o desenvolvimento tecnológico ocorra com equilíbrio entre os interesses econômicos e a proteção de direitos fundamentais, garantindo que a inteligência artificial seja uma aliada no fortalecimento da democracia, da justiça social e do bem comum.

Diante do avanço acelerado da inteligência artificial, torna-se imprescindível que governos, empresas, pesquisadores e a sociedade civil atuem de forma articulada na construção de marcos regulatórios e estruturas de governança que priorizem a ética, a transparência e os direitos fundamentais.



A IA, por sua complexidade e alcance, exige um olhar atento, crítico e, acima de tudo, responsável. Promover a inovação com equilíbrio, assegurando que seus benefícios sejam amplamente distribuídos e que seus riscos sejam devidamente mitigados, é o desafio central do nosso tempo.

A construção de um ecossistema de inteligência artificial justo e confiável não é tarefa de um único ator, mas de um esforço coletivo e contínuo, em que a governança se apresenta como instrumento importante para orientar o presente e projetar um futuro mais ético, inclusivo e sustentável.

REFERÊNCIAS

AUTRAN, F. IA da Amazon usada em análise de currículos discriminava mulheres. 2018. Disponível em: <https://www.tecmundo.com.br/software/135062-ia-amazon-usada-analise-curriculos-discriminava-mulheres.htm?ab=true&>. Acesso em: 09 abr. 2025.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet). Diário Oficial da União: seção 1, Brasília, DF, p. 1, 15 ago. 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 09 abr. 2025.

BRASIL. **Lei nº 9.610, de 19 de fevereiro de 1998**. Altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências. Diário Oficial da União: seção 1, Brasília, DF, p. 3, 20 fev. 1998. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l9610.htm. Acesso em: 09 abr. 2025.

BRASIL. Senado Federal. **Projeto de Lei nº 2.338, de 2023**. Estabelece princípios, direitos e deveres para o uso da inteligência artificial no Brasil e cria o Sistema Nacional de Regulação e Governança de Inteligência Artificial. Brasília, DF, 2023. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/164529>. Acesso em: 09 abr. 2025.

CAO, C.; REUTERS. 2025. **ChatGPT image tool draws a million new users in one hour**. In: The Standard. Disponível em: <https://www.thestandard.com.hk/section-news/section/2/271420/ChatGPT-image-tool-draws-a-million-new-users-in-one-hour>. Acesso em: 09 abr. 2025.

HARADA, E. Y. **Fail épico**: sistema do Google Fotos identifica pessoas negras como gorilas. 2015. Disponível em: <https://www.tecmundo.com.br/google-fotos/82458-polemica-sistema-google-fotos-identifica-pessoas-negras-gorilas.htm?ab=true&>. Acesso em: 09 abr. 2025.

FRAZÃO, A.; MULHOLLAND, C. (Coord.). **Inteligência Artificial e Direito**: ética, regulação e responsabilidade. São Paulo: Thomson Reuters Brasil, 2019.

MAZZALI, R.; ERCOLIN, C. A. **Governança corporativa**. 1. ed. Rio de Janeiro: Editora FGV, 2018.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (OCDE). **Princípios da OCDE sobre Inteligência Artificial**. 2024. Disponível em: <https://www.oecd.org/en/topics/ai-principles.html>. Acesso em: 09 abr. 2025.

O'NEIL, C. **Algoritmos de Destruição em Massa**: Como o Big Data Aumenta a Desigualdade e Ameaça a Democracia. Tradução de Rafael Abraham. 1. ed. Santo André: Rua do Sabão, 2020.

LARSON, J.; MATTU, S.; KIRCHNER, L.; ANGWIN, J. **How We Analyzed the Recidivism Algorithm**. 2016. Disponível em: <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>. Acesso em: 09 abr. 2025.

PINHEIRO, C. O ChatGPT é preconceituoso? Reflexões em torno dos sistemas de inteligência artificial e seus vieses à luz da Lei Geral de Proteção de Dados. In: XIII ENCONTRO INTERNACIONAL DO CONPEDI – URUGUAI: **Governo Digital, Direito e Novas Tecnologias**, 2024, Montevideu. Anais do XIII Encontro Internacional do CONPEDI – Uruguai. Florianópolis: CONPEDI, 2024. ISBN 978-85-5505-986-5. Disponível em: www.conpedi.org.br. Acesso em: 09 abr. 2025.

NDTV NEWS DESK. **Insult To Life Itself**: Ghibli Founder Hayao Miyazaki On AI-Generated Art. 2025. Disponível em: https://www.ndtv.com/world-news/quot-i-would-never-incorporate-this-quot-what-studio-ghibli-039-s-hayao-miyazaki-once-said-about-ai-animation-8021037?utm_source=chatgpt.com. Acesso em: 09 abr. 2025.

UNESCO. **Recomendação sobre a Ética da Inteligência Artificial**. 2021. Disponível em: <https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence>. Acesso em: 09 abr. 2025.

UNIÃO EUROPEIA. Parlamento Europeu. **Regulamento (UE) do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de inteligência artificial (Lei de Inteligência Artificial)**. 2024. Disponível em: <https://www.europarl.europa.eu/news/pt/press-room/20240308IPR19022/parlamento-aprova-primeira-lei-da-ue-sobre-inteligencia-artificial>. Acesso em: 09 abr. 2025.

WARDLE, C.; DERAKHSHAN, H. **Information disorder**: Toward an interdisciplinary framework for research and policymaking. Council of Europe, 2017. Disponível em: <https://www.coe.int/en/web/freedom-expression/information-disorder>. Acesso em: 14 jun. 2024.

WORLD ECONOMIC FORUM. **Artificial Intelligence for Efficiency, Sustainability and Inclusivity**. 2025. Disponível em: <https://www.weforum.org/publications/artificial-intelligence-for-efficiency-sustainability-and-inclusivity-in-tradetech>. Acesso em: 09 abr. 2025.

SENAI <LAB365>