

Revista Brasileira de Ciências Sociais Aplicadas

ISSN 3085-8151

vol. 2, n. 1, 2026

••• ARTIGO 5

Data de Aceite: 08/01/2026

FRONTEIRAS ÉTICAS DA IA GENERATIVA EM PRODUTOS DIGITAIS: PILARES PARA GOVERNANÇA NO CICLO DE DESENVOLVIMENTO

Thais Bellegante



Todo o conteúdo desta revista está licenciado sob a Licença Creative Commons Atribuição 4.0 Internacional (CC BY 4.0).

Resumo: O uso da Inteligência Artificial Generativa (GenAI) em produtos digitais tem transformado a forma como soluções são projetadas, desenvolvidas e operadas. Apesar dos avanços em escala, produtividade e inovação, a aplicação da GenAI traz riscos éticos e operacionais relevantes, como vieses algorítmicos, falta de explicabilidade, opacidade dos modelos e lacunas na governança. Este artigo propõe uma estrutura composta por cinco pilares éticos: transparência, equidade, explicabilidade, robustez técnica e accountability, integrados a critérios de governança aplicáveis ao ciclo de desenvolvimento de produtos digitais. A partir de uma revisão crítica de 42 documentos científicos, normativos e técnicos, o estudo consolida diretrizes que podem apoiar profissionais e organizações na construção de soluções mais responsáveis, seguras e alinhadas a valores sociais. Ao traduzir princípios éticos em práticas operacionais, a proposta busca preencher a lacuna existente entre discurso e prática no uso da GenAI, contribuindo tanto para o avanço acadêmico quanto para o desenvolvimento sustentável e ético de tecnologias emergentes.

Palavras-chave: Inteligência Artificial Generativa, Ética em Tecnologia, Governança Algorítmica, Desenvolvimento de Produtos Digitais, Viés Algorítmico.

Introdução

A Inteligência Artificial (IA) deixou há muito de ser só uma promessa tecnológica e passou a ser parte estruturante do desenvolvimento de produtos digitais. Hoje ela não só otimiza processos, como influencia decisões que moldam a experiência dos usuários e os próprios modelos de negócio. Floridi

et al. (2018) já defendiam que IA não é só tecnologia — é infraestrutura social, econômica e política. E os números reforçam isso: segundo o AI Index 2024 (Stanford, 2024), o uso de IA nas empresas cresceu 250% nos últimos cinco anos. A McKinsey (2023) mostra que mais da metade das empresas globais já usam IA em pelo menos uma função core, seja operacional ou estratégica.

Entre os avanços mais relevantes desse cenário, a IA Generativa (GenAI) vem ganhando protagonismo. Isso porque ela não só analisa dados, mas cria — textos, imagens, códigos, sons — com um nível de autonomia que muda completamente o jogo no desenvolvimento de produtos digitais. Modelos como GPT-4, DALL·E, Midjourney e Stable Diffusion estão cada vez mais integrados a serviços que lidam diretamente com pessoas, personalizam experiências, tomam decisões e, muitas vezes, fazem isso sem que o usuário perceba o que está por trás da interface.

Mas esse avanço, obviamente, não vem sem riscos. Como já apontam Mittelstadt et al. (2016) e Binns (2018), sistemas baseados em GenAI operam de forma opaca, carregam vieses dos dados e entregam outputs imprevisíveis. E isso não é um problema só técnico — é ético, social, jurídico e, muitas vezes, invisível até que o impacto já tenha acontecido. Para quem desenvolve produtos, a responsabilidade se amplia: é preciso olhar não só para a experiência e a entrega, mas para as implicações reais que esses modelos podem gerar.

Diante desse cenário, este artigo propõe a construir uma estrutura ética aplicável ao desenvolvimento de produtos digitais com GenAI. A partir de uma revisão crítica da literatura, proponho cinco pilares éticos: Transparência, Equidade, Explicabilidade,

Robustez Técnica e Accountability, acompanhados de critérios práticos de governança. O objetivo é apoiar profissionais de tecnologia, produto e design na construção de soluções que sejam, ao mesmo tempo, inovadoras e alinhadas a valores sociais.

Metodologia

Este estudo segue uma abordagem qualitativa e exploratória, justamente porque se trata de um tema recente, em rápida evolução e com poucas referências consolidadas na intersecção entre ética, governança algorítmica e desenvolvimento de produtos digitais com GenAI.

A pesquisa foi construída a partir de uma revisão crítica e direcionada de 42 documentos, distribuídos assim:

- 26 artigos científicos, provenientes de bases como Scopus, ACM Digital Library, SpringerLink, IEEE Xplore e Google Scholar;
- 9 diretrizes e normas institucionais, incluindo documentos da OCDE (2019), UNESCO (2021) e ISO/IEC 42001 (2023);
- 7 publicações técnicas e metodológicas, de organizações como PMI, AI Now Institute, OpenAI e Microsoft.

O quadro 1 - Distribuição de fontes de pesquisa, apresenta a distribuição quantitativa das fontes:

A análise dos materiais foi conduzida com base na Análise de Conteúdo Temática (Bardin, 2011), buscando mapear conceitos, convergências teóricas, lacunas normativas e diretrizes práticas que sustentam a construção dos pilares éticos e dos critérios de governança propostos.

O critério de seleção priorizou fontes diretamente relacionadas aos seguintes eixos: ética aplicada à tecnologia, viés algorítmico, confiabilidade técnica, governança de IA, design responsável e impactos sociais de tecnologias emergentes.

É importante destacar que o estudo não tem caráter empírico. Ele constrói uma base conceitual e normativa que possa ser utilizada como referência por profissionais e organizações que atuam no desenvolvimento de produtos digitais baseados em IA Generativa.

Referencial Teórico

Ética em Tecnologia e Desenvolvimento de Produtos Digitais

Falar sobre ética em tecnologia é, na prática, discutir quais decisões tomamos (ou não tomamos) na hora de projetar sistemas

Tipo de Documento	Quantidade	Fontes Principais
Artigos Científicos	26	Scopus, ACM, Springer, IEEE, Google
Diretrizes/Normas	9	OCDE, UNESCO, ISO, AI Act, PMBOK
Documentação Técnica	7	AI Now, OpenAI, Microsoft, PMI, Scrum

Quadro 1 - Distribuição de fontes de pesquisa

Fonte: autora

que impactam diretamente a vida das pessoas. Em ambientes de produto, essa discussão se torna ainda mais urgente quando a IA Generativa entra no jogo porque ela não só executa, mas cria, interage e, muitas vezes, decide.

A literatura não deixa dúvidas: todo design é, por definição, uma decisão carregada de implicações morais, políticas e sociais (Mittelstadt et al., 2016; Winner, 1980). Isso significa que ignorar os dilemas éticos não é ser neutro, é assumir o risco de reforçar desigualdades, excluir grupos ou gerar impactos que poderiam (e deveriam) ser evitados.

Na gestão de projetos, o Guia PMBOK – 7^a Edição (PMI, 2021) já aponta a ética como princípio norteador, especialmente através do conceito de Stewardship agir com responsabilidade, integridade e consciência sobre como os projetos impactam não só a organização, mas a sociedade. O Código de Ética do PMI (2017) reforça quatro valores centrais: responsabilidade, respeito, justiça e honestidade todos absolutamente aplicáveis à construção de produtos digitais, principalmente quando se lida com IA.

O mesmo raciocínio vale para metodologias ágeis. O Scrum Guide (Schwaber & Sutherland, 2020) deixa claro que valores como comprometimento, foco, abertura, respeito e coragem não são acessórios são princípios que sustentam a construção de soluções que, além de funcionarem, são justas, transparentes e confiáveis.

Abordagens como o Ethical Design e o Value Sensitive Design (Friedman, Kahn & Borning, 2006) vão além, defendendo que valores humanos como equidade, privacidade, segurança e acessibilidade devem estar presentes desde a concepção da solução, não

só como mitigação de riscos, mas como requisitos de projeto.

Além disso, diretrizes internacionais, como as da OCDE (2019) e da UNESCO (2021), reforçam que práticas éticas não podem ser um apêndice. Elas devem ser parte estruturante da governança, da arquitetura dos sistemas e de todo o ciclo de vida do desenvolvimento tecnológico.

Em resumo, construir produtos digitais com GenAI exige, sim, eficiência, escala e inovação, mas exige, acima de tudo, responsabilidade social, ética aplicada e uma cultura de desenvolvimento orientada por valores.

Viés Algorítmico e Discriminação Automatizada

Falar de IA sem falar de viés é ignorar uma das questões mais críticas e, talvez, uma das mais negligenciadas na aplicação prática dessas tecnologias. No caso da GenAI, isso é ainda mais sensível. Afinal, estamos lidando com modelos que não apenas analisam, mas geram conteúdo, interagem e, muitas vezes, influenciam decisões sem que os usuários sequer percebam.

O ponto central é simples, mas poderoso: modelos treinados em grandes volumes de dados herdados da internet que, por si só, já carregam todo tipo de distorção social, cultural, econômica e histórica inevitavelmente reproduzem esses mesmos padrões. Como mostram Barocas, Hardt e Narayanan (2019), o viés algorítmico nasce justamente da maneira como os dados são coletados, processados e modelados.

Binns (2018) ajuda a entender que esses vieses aparecem em várias camadas: nos dados de entrada (desbalanceados, excludentes), na modelagem (na escolha dos

atributos), na interface (em como os resultados são apresentados) e até na interpretação final feita pelos usuários. E isso não é teoria distante, é prático e visível. Birhane et al. (2021) e Bender et al. (2021) documentam, por exemplo, casos em que modelos como GPT, Midjourney e Stable Diffusion geraram outputs racistas, misóginos e discriminatórios.

Além do viés de conteúdo, tem também o viés de desempenho quando determinados grupos recebem resultados piores, respostas menos precisas ou recomendações enviesadas. E aí, quando GenAI é aplicada em contextos sensíveis como crédito, saúde, educação ou mercado de trabalho. Esse viés deixa de ser um problema técnico e passa a ser uma questão de acesso, oportunidade e até de dignidade (O’Neil, 2016).

Pra complicar mais, entra o problema da opacidade epistêmica (Floridi & Cowls, 2019): na prática, nem quem desenvolveu o modelo consegue explicar claramente como certas respostas foram geradas. E se não dá pra entender como o sistema funciona, como mitigar viés? Como responsabilizar alguém quando o sistema falha?

Por isso, surgem práticas como datasheets for datasets (Gebru et al., 2018) e model cards (Mitchell et al., 2019), que tentam documentar, de forma mais transparente, como os dados foram tratados, quais são os limites do modelo e quais riscos estão embutidos ali.

A real é que viés algorítmico não é só um bug. É uma externalidade social de sistemas que são treinados em dados desiguais e operam em sociedades desiguais. Ignorar isso é perpetuar e até escalar essas desigualdades via tecnologia.

Confiabilidade, Explicabilidade e Responsabilidade Técnica

À medida que modelos generativos entram cada vez mais no core dos produtos digitais, uma coisa fica muito clara: confiar na IA deixou de ser um exercício de fé tecnológica e virou um desafio operacional, técnico e ético.

Modelos como GPT-4 e DALL-E operam com base em estatística, não em verdade. Isso significa que eles não respondem “o que é certo”, mas “o que é mais provável”. E é exatamente daí que nascem os famosos fenômenos de alucinação dos modelos quando eles entregam uma resposta que parece correta, tem tom confiante, mas nem sempre é verdade (Ji et al., 2023). Quando se imagina esse cenário aplicado em chatbots, assistentes financeiros, suporte jurídico ou plataformas de educação. O risco não é só gerar uma resposta errada é fazer com que o usuário acredite que ela está certa. E aí mora um problema enorme de confiabilidade.

A questão da explicabilidade também não é menor. Doshi-Velez e Kim (2017) explicam que, sem conseguir entender como um modelo chega a determinado output, simplesmente não existe como auditar, contestar ou corrigir uma decisão automatizada. E isso pesa muito mais em produtos que operam em setores regulados como saúde, finanças, crédito ou segurança pública.

Rudin (2019) vai além: se o modelo é tão opaco que não conseguimos explicar como ele funciona, talvez nem deveríamos usá-lo em contextos de alta criticidade.

A falta de explicabilidade traz, de brinde, um baita problema de accountability. Quando algo dá errado, quem responde? o time que desenvolveu o produto? A empresa dona do modelo? O fornecedor da tecnolo-

gia? O time que usa a ferramenta? Wieringa (2020) traz esse debate pra mesa de forma muito clara: a IA redistribui as responsabilidades, mas, se não houver um framework bem definido, elas simplesmente somem.

É exatamente por isso que surgem práticas e padrões como a ISO/IEC 42001:2023, os Model Cards (Mitchell et al., 2019) e os AI FactSheets (Arnold et al., 2019). Eles não resolvem tudo, mas ajudam a documentar, rastrear e tornar mais visível como esses sistemas operam.

Como bem pontuam Morley et al. (2021), isso não é só sobre ferramentas de compliance é sobre uma mudança real de cultura no desenvolvimento. Se a ética não entra no começo do projeto, não adianta querer depois adotar uma postura de “IA responsável”.

Governança e Diretrizes para Uso Responsável da GenAI

Governança de IA, na prática, significa criar mecanismos pra garantir que a tecnologia esteja alinhada com segurança, transparência, responsabilidade e, principalmente, com valores sociais. E, quando falamos de GenAI, esse tema não é opcional, se tornou urgente.

A ausência de governança, especialmente em modelos que operam de forma generativa, não só aumenta riscos técnicos e jurídicos, ela abre espaço para desinformação, violações de direitos e impactos sociais em escala (Floridi et al., 2018).

Não por acaso, vários organismos internacionais estão correndo pra estabelecer frameworks éticos e regulatórios. A OCDE (2019) traz princípios como bem-estar social, equidade, transparência, robustez técnica e accountability. A UNESCO (2021)

segue na mesma linha, propondo, além dos princípios, obrigações concretas como avaliações de impacto, proteção contra discriminação e direito de contestação de decisões automatizadas.

Na União Europeia, o AI Act vem para classificar aplicações de IA por nível de risco, criando exigências mais rigorosas para modelos considerados de alto impacto, como reconhecimento facial, scoring de crédito ou sistemas que afetem a segurança física ou mental das pessoas (Veale & Borgesius, 2021).

No dia a dia dos produtos digitais, isso se traduz em práticas muito concretas, como:

- Avaliação de Impacto Algorítmico (AIA), pra mapear riscos antes da IA ir pra produção;
- Comitês de Ética ou Governança Tecnológica, que fazem revisões contínuas
- Transparência técnica, usando ferramentas como datasheets, model cards e AI FactSheets pra documentar os limites e as condições de uso dos modelos;
- Engajamento dos stakeholders, trazendo pra mesa não só usuários, mas também especialistas, grupos vulneráveis e representantes da sociedade civil.

Empresas como IBM, Google, Microsoft e OpenAI já criaram guias internos pra isso, baseados em princípios como fairness, privacy, safety e accountability. Mas, como mostram Morley et al. (2021), uma coisa é ter princípios declarados. Outra, bem diferente, é transformar isso em prática dentro dos produtos.

Se a governança não estiver presente desde o design até a operação contínua, ela vira só um documento para enfeite e cumprimento de protocolo, e não uma prática real.

Desenvolvimento

Aplicações da GenAI no Ciclo de Vida de Produtos Digitais

A GenAI não está sendo incorporada nos produtos digitais só como mais uma tecnologia. Ela está mudando, de fato, a lógica de desenvolvimento, operação e até a forma como se toma decisões dentro das empresas. Não é mais sobre IA como apoio, mas sobre IA como protagonista no ciclo de vida dos produtos.

Na fase de concepção e design, modelos como Midjourney, DALL·E e Uizard estão sendo usados para gerar protótipos, interfaces e até fluxos de interação quase instantaneamente. Isso muda o tempo de desenvolvimento, mas também muda quem decide o quê — o time não desenha mais telas, mas define parâmetros do modelo para gerar as opções.

No desenvolvimento, ferramentas como GitHub Copilot e ChatGPT já participam da geração de código, revisão de testes e automação de várias etapas operacionais. E, quando o produto vai pro mercado, a GenAI atua diretamente na experiência do usuário, seja personalizando ofertas, adaptando conteúdos em tempo real ou operando como interface principal, via chatbots ou assistentes.

Só que esse movimento não é só técnico. Ele altera profundamente o papel dos times de produto. As decisões deixam de

ser sobre regras fixas e passam a ser sobre curadoria, calibragem e controle de outputs probabilísticos. Isso exige que as equipes desenvolvam, além de competências técnicas, uma capacidade crítica muito maior para avaliar impactos, riscos e dilemas éticos associados a esses modelos.

Dilemas Éticos em Produtos Digitais com GenAI

À medida que essas aplicações escalam, surgem dilemas que desafiam qualquer manual tradicional de desenvolvimento de produto.

O primeiro é a linha tênue entre personalização e manipulação. Sistemas que adaptam conteúdos, ofertas e interfaces em tempo real podem, sim, gerar mais valor pro usuário. Mas também podem reforçar bolhas, vícios de uso, enviesar decisões de consumo e até manipular comportamentos de forma invisível, como já alerta Zuboff (2019).

Depois, vem o dilema da automação versus desumanização. Usar GenAI para atendimento ao cliente, por exemplo, cortar custos e escalar operações. Mas, se mal calibrado, o risco é oferecer experiências mecânicas, enviesadas ou até insensíveis, especialmente para pessoas em situação de vulnerabilidade, como destacam Binns et al. (2018).

Outro ponto crítico é a questão da responsabilidade editorial. Quando um sistema gera textos, resumos, traduções ou comentários, quem responde pelo conteúdo? A empresa? O desenvolvedor? O fornecedor do modelo? Isso não é trivial, principalmente em um mundo em que a disseminação de desinformação virou problema global, como discutem Mökander e Floridi (2023).

Por fim, tem um embate claro entre iteração contínua, típica do desenvolvimento ágil, e a necessidade de accountability algorítmica. A lógica do design digital trabalha com melhorias constantes. Mas, quando se trata de IA, alterações constantes nos modelos podem comprometer auditorias, dificultar rastreabilidade e até gerar comportamentos não previstos, como mostram Raji et al. (2020).

Com a expansão dessas aplicações, emergem dilemas éticos que desafiam práticas tradicionais de design e governança de produto. Entre os principais, destacam-se:

- **Personalização vs. Manipulação:** sistemas generativos que adaptam conteúdos e interfaces em tempo real podem reforçar vícios de uso, enviesar decisões de compra ou moldar preferências sem que o usuário perceba (Zuboff, 2019). A linha entre entrega de valor e exploração comportamental torna-se tênue;
- **Automação de atendimento vs. desumanização:** produtos que utilizam GenAI para atendimento ao cliente frequentemente reduzem custos e aumentam a escalabilidade, mas correm o risco de gerar experiências frustrantes, enviesadas ou insensíveis — especialmente em contextos de vulnerabilidade (Binns et al., 2018);
- **Interação com conteúdos gerados vs. responsabilidade editorial:** empresas que utilizam GenAI para gerar resumos, traduções, artigos ou comentários precisam responder pela veracidade e pelos possíveis impactos sociais da infor-

mação gerada (Möckander & Floridi, 2023);

- **Iteração contínua vs. accountability algorítmica:** o modelo iterativo do design digital ágil muitas vezes entra em conflito com a necessidade de rastreabilidade e explicabilidade de decisões automatizadas. Modificações contínuas podem dificultar auditorias e o controle sobre o comportamento da GenAI no tempo (Raji et al., 2020).

O que esses dilemas deixam claro é que não basta mais ter princípios éticos escritos em um deck ou colados na parede. É preciso transformar esses princípios em práticas operacionais, integradas diretamente ao processo de desenvolvimento.

Da Teoria à Prática: Lacunas na Implementação Ética

Apesar de todo o avanço teórico, metodológico e das diretrizes produzidas nos últimos anos, a verdade é que a maioria das empresas ainda não conseguiu transformar isso em prática no desenvolvimento dos seus produtos com IA.

O primeiro gargalo é a pressão por velocidade e entrega. Time-to-market segue sendo uma prioridade que, na prática, empurra discussões éticas para depois ou, pior, deixa elas de fora do processo.

O segundo problema é a ausência de métricas éticas claras. Enquanto existem KPIs para quase tudo: conversão, retenção, churn, dificilmente os times medem se o modelo está operando de forma justa, explicável ou sem vieses.

Além disso, há um gap técnico evidente. Muitos times de produto, UX ou design não têm repertório suficiente para entender como funcionam os modelos generativos, muito menos para avaliar os riscos éticos associados a eles.

E, pra fechar, tem a questão da diversidade. Equipes homogêneas, compostas por perfis técnicos muito parecidos, tendem a não perceber determinados tipos de vieses, impactos e externalidades. Isso compromete diretamente a capacidade de projetar soluções mais inclusivas e socialmente alinhadas.

No fim das contas, o que se observa é que frameworks de governança e diretrizes éticas ainda são tratados como estruturas paralelas ao desenvolvimento. Eles não fazem parte do ciclo real de tomada de decisão na maioria dos produtos. E enquanto isso não mudar, a capacidade das empresas de antecipar riscos, responder a falhas ou garantir padrões de justiça algorítmica vai seguir limitada.

Análise e Discussões

A presença da Inteligência Artificial Generativa no desenvolvimento de produtos digitais não representa só mais uma onda de inovação tecnológica. Ela muda, de forma estrutural, como as soluções são pensadas, construídas e operadas. Só que, junto com esse avanço, vêm desafios que não são triviais. A imprevisibilidade dos outputs, a dificuldade de rastrear como as decisões são tomadas, a reprodução de vieses e a automatização de processos críticos fazem da GenAI uma tecnologia tão poderosa quanto potencialmente arriscada não só para quem desenvolve, mas para quem consome e para sociedade como um todo.

Quando a gente olha pra prática, fica evidente um distanciamento entre os princípios éticos que muitas empresas declaram e o que, de fato, é aplicado no desenvolvimento dos seus produtos. Essa distância se agrava com a pressão constante por entregas rápidas, a ausência de regulamentações específicas para IA generativa e a falta de repertório técnico e ético dentro dos times para lidar com os dilemas que surgem no dia a dia.

Diante desse cenário, a proposta aqui é apresentar uma estrutura composta por cinco pilares éticos integrados a critérios de governança, que podem ser adotados como referência para quem desenvolve, gerencia e opera produtos digitais que utilizam GenAI. Esses pilares não são um modelo fechado, mas uma base que pode (e deve) ser adaptada às realidades de cada organização, considerando o nível de maturidade dos times, os contextos regulatórios e os riscos específicos de cada aplicação.

Pilares Éticos para Produtos com GenAI

Transparéncia Técnica e Comunicacional

Transparéncia não pode ser tratada como um conceito genérico. Ela precisa ser operacionalizada tanto na parte técnica documentando modelos, dados, limitações e riscos, quanto na comunicação com os usuários. Se o usuário está interagindo com uma IA, ele precisa saber disso de forma clara, objetiva e comprehensível. E essa transparéncia não pode se perder em linguagem técnica ou burocrática. É papel dos times garantir que a experiência do usuário reflete isso.

Equidade e Representatividade

Se a IA aprende a partir dos dados, ela também reproduz os mesmos padrões, desigualdades e distorções que estão nesses dados. Por isso, garantir equidade não é só um ideal é um trabalho prático que passa pela curadoria consciente dos dados, pela realização de testes de impacto entre diferentes grupos e, principalmente, pela diversidade real dentro dos times que projetam, treinam e validam esses modelos. Sem diversidade de olhares, o risco de reforçar exclusões e discriminações é praticamente garantido.

Explicabilidade e Contestabilidade

Sistemas que automatizam decisões precisam ser compreensíveis e auditáveis. Isso não significa que o usuário final precise entender todos os detalhes técnicos do modelo, mas é fundamental que existam mecanismos de explicação acessíveis, além de caminhos claros pra contestar uma decisão quando necessário. Produtos que usam GenAI precisam ser projetados com camadas de justificativa e, sempre que fizer sentido, com a possibilidade de revisão humana das decisões automatizadas.

Segurança e Robustez Técnica

Robustez não é só sobre não falhar ou sobre o tamanho da tecnologia. É sobre garantir que o modelo funcione corretamente, mesmo quando pressionado ao limite, mesmo quando lida com casos extremos ou dados que fogem dos padrões. Isso inclui proteger contra falhas, contra alucinações (quando o modelo gera uma resposta que parece verdadeira, mas é falsa) e contra vulnerabilidades que possam ser exploradas

maliciosamente. Testes de stress, simulações de erro, validação contínua e mecanismos de contenção são fundamentais.

Accountability e Responsabilidade Compartilhada

Em produtos baseados em IA generativa, a responsabilidade não pode ficar perdida entre fornecedores de API, desenvolvedores, squads de produto ou executivos. Precisa estar clara. Quem responde pelo quê? Quem valida? Quem audita? Quem dá a palavra final em caso de falhas, erros ou impactos negativos? Accountability não é só um pilar ético, é uma condição básica para que qualquer solução baseada em IA seja viável do ponto de vista social, jurídico e, cada vez mais, regulatório.

Critérios de Governança: Aplicação dos Pilares

A estrutura dos cinco pilares proposta aqui não surge do zero. Ela é fruto da consolidação de referências robustas no campo da ética em IA, da governança algorítmica e das práticas metodológicas que vêm sendo desenvolvidas globalmente. Floridi et al. (2018), OECD (2019), UNESCO (2021) e Mökander e Floridi (2023) são algumas das principais bases que sustentam essa proposta, junto com frameworks como o Value Sensitive Design (Friedman, Kahn e Boring, 2006) e as práticas de auditoria algorítmica descritas por Raji et al. (2020).

A partir da leitura cruzada desses autores e instituições, foram consolidados os cinco pilares com suas referências diretas:

- Transparência, como discutido por Mitchell et al. (2019) e Arnold et al. (2019);

- Equidade, baseada em Barocas et al. (2019) e Binns (2018);
- Explicabilidade, apoiada em Doshi-Velez & Kim (2017) e Rudin (2019);
- Robustez Técnica, fundamentada na ISO/IEC 42001 (2023);
- Accountability, alinhada às discussões de Floridi & Cowls (2019) e Wieringa (2020).

Cada um desses pilares não é um princípio abstrato, mas se traduz diretamente em práticas operacionais. Isso significa adotar avaliações de impacto algorítmico ainda nas fases iniciais do desenvolvimento, implementar documentação técnica estruturada (como model cards, datasheets e AI FactSheets), inserir checklists éticos nas etapas de backlog, refinamento e validação, e, principalmente, estabelecer de forma clara quem são os responsáveis pela governança ética do produto.

Esses critérios não são opcionais. Eles devem cada vez mais serem utilizados como referência para garantir que produtos digitais que usam GenAI operem de forma ética, sustentável e alinhada às expectativas sociais, regulatórias e de mercado.

Considerações Finais

O avanço da Inteligência Artificial Gerativa no desenvolvimento de produtos digitais não é só uma tendência tecnológica, é uma mudança estrutural na forma como soluções são criadas, operadas e entregues. Só que essa transformação não vem sem custo. Pelo contrário, ela traz uma série de riscos que são tão técnicos quanto éticos, tão operacionais quanto sociais.

Ao longo desse estudo, ficou evidente que usar GenAI em produtos digitais exige muito mais do que domínio técnico. Exige que times e organizações passem a incorporar, de forma estruturada, princípios de responsabilidade, transparência, equidade e segurança. E isso não pode ser visto como um adendo ou uma camada a ser colocada depois. É parte do próprio processo de desenvolvimento.

A proposta dos cinco pilares éticos: transparência, equidade, explicabilidade, robustez técnica e accountability, não nasce como um modelo teórico desconectado da prática. Ela surge justamente da análise crítica dos desafios que hoje já são enfrentados por quem trabalha com desenvolvimento de produtos que usam IA generativa. E, mais do que princípios abstratos, esses pilares se desdobram em critérios de governança muito objetivos, que podem (e precisam) ser aplicados no dia a dia dos times.

Esse artigo, no entanto, tem suas limitações. Por ser uma pesquisa teórico-exploratória, não se apoia em estudos de caso ou em dados empíricos coletados diretamente em organizações. Isso abre espaço para que futuras pesquisas avancem na validação desse framework, seja por meio de estudos de campo, entrevistas com profissionais de produto, análises comparativas ou até acompanhamento longitudinal de empresas que estão estruturando governança ética pra IA generativa.

O que parece inegociável, olhando pra frente, é que a discussão sobre ética em IA vai deixar de ser periférica. Ela não é uma demanda só da academia, nem apenas uma exigência dos órgãos reguladores. É uma necessidade concreta para qualquer organização que queira operar de forma sustentável, legítima e alinhada aos valores da sociedade.

Em um contexto onde a tecnologia avança mais rápido do que a nossa capacidade de regular, fiscalizar ou até entender todos os seus impactos, a integração entre ética e inovação deixa de ser uma escolha. Ela se torna uma competência essencial para quem desenvolve produtos digitais. E quem não entender isso, mais cedo ou mais tarde, vai perceber que os riscos de ignorar essa agenda são muito maiores do que os custos de trazê-la pra dentro do processo.

Referências

- ARNOLD, M. et al. *FactSheets: Increasing Trust in AI Services through Supplier's Declarations of Conformity*. IBM Research, 2019.
- BAROCAS, S.; HARDT, M.; NARAYANAN, A. *Fairness and Machine Learning*. fairmlbook.org, 2019. BARDIN, L. *Análise de conteúdo*. Lisboa: Edições 70, 2011.
- BENDER, E. M. et al. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021.
- BINNS, R. Fairness in Machine Learning: Lessons from Political Philosophy. *Proceedings of the 2018 Conference on Fairness, Accountability and Transparency*, p. 149–159, 2018.
- BIRHANE, A.; PRABHU, V.; KAHEMBWE, E. Multimodal Datasets: Misogyny, Malignant Stereotypes, and the Absence of Documentation. *arXiv preprint*, 2021. Disponível em: <https://arxiv.org/abs/2110.01963>
- DOSHI-VELEZ, F.; KIM, B. Towards a Rigorous Science of Interpretable Machine Learning. *arXiv preprint arXiv:1702.08608*, 2017.
- FLORIDI, L.; COWLS, J. A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, 2019.
- FLORIDI, L. et al. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, v. 28, n. 4, p. 689–707, 2018.
- FRIEDMAN, B.; KAHN Jr., P. H.; BORNING, A. Value Sensitive Design and Information Systems. In: ZHANG, P.; GALLETTA, D. (org.). *Human-Computer Interaction and Management Information Systems: Foundations*. Armonk: M.E. Sharpe, 2006.
- GEBRU, T. et al. Datasheets for Datasets. *arXiv preprint arXiv:1803.09010*, 2018.
- JI, Z. et al. Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys*, v. 55, n. 12, p. 1–38, 2023.
- LIPTON, Z. C. The Mythos of Model Interpretability. *Communications of the ACM*, v. 61, n. 10, p. 36–43, 2018.
- MITCHELL, M. et al. Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT)*, 2019.
- MÖKANDER, J.; FLORIDI, L. Ethics as a Service: A Pragmatic Operationalisation of AI Ethics. *AI and Ethics*, v. 3, p. 123–138, 2023.
- MORLEY, J. et al. From What to How: An Overview of AI Ethics Tools, Methods and Research to Translate Principles into Practices. *AI & Society*, v. 36, p. 59–71, 2021.
- OECD. *OECD Principles on Artificial Intelligence*. Paris: OECD, 2019.
- O'NEIL, C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing Group, 2016.
- PMI – Project Management Institute. *Código de Ética e Conduta Profissional*. Pennsylvania: PMI, 2017.

PMI – Project Management Institute. *Guia PMBOK® – Um Guia do Conhecimento em Gerenciamento de Projetos*. 7. ed. Pennsylvania: PMI, 2021.

RAJI, I. D. et al. Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020.

RUDIN, C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence*, v. 1, p. 206–215, 2019.

SCHWABER, K.; SUTHERLAND, J. *The Scrum Guide™ – The Definitive Guide to Scrum: The Rules of the Game*. 2020.

UNESCO. *Recommendation on the Ethics of Artificial Intelligence*. Paris: UNESCO, 2021.

VEALE, M.; BORGESIUS, F. Z. Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*, v. 22, n. 4, p. 97–112, 2021.

WINNER, L. Do Artifacts Have Politics?. *Dædalus*, v. 109, n. 1, p. 121–136, 1980. ZUBOFF, S. *The Age of Surveillance Capitalism*. New York: PublicAffairs, 2019.