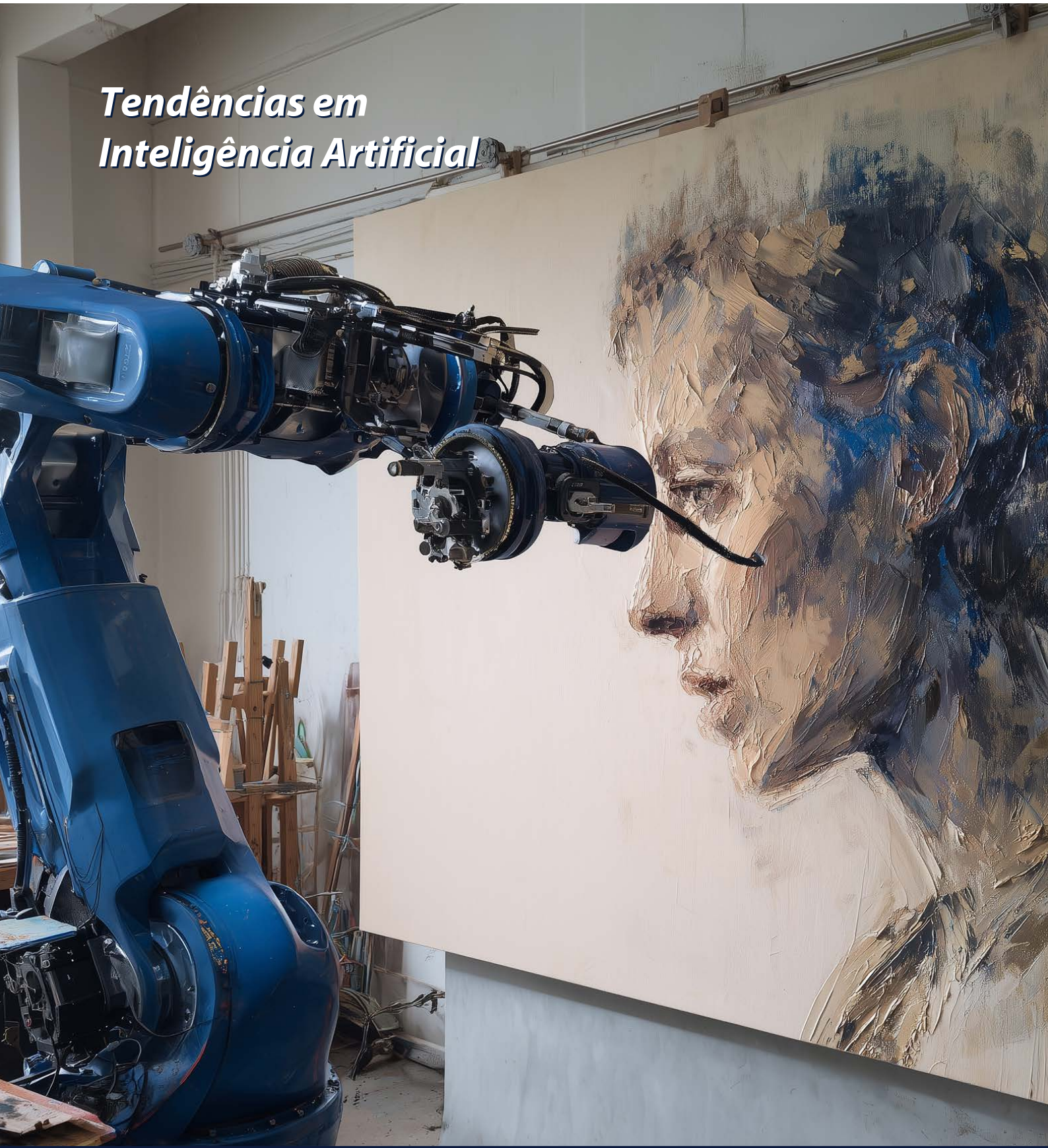
Management Solutions
Making things happen

Tendências em Inteligência Artificial



www.managementsolutions.com

Design e diagramação
Departamento de Marketing e Comunicação - Management Solutions

Fotografias
Arquivo fotográfico da Management Solutions (algumas imagens geradas com o uso de IA)
AdobeStock

Sumário

01 | Introdução

"Se em 2020 te disséssemos que estaríamos onde estamos hoje, provavelmente pareceria mais louco do que nossas previsões atuais para 2030".

Sam Altman[1]

A inteligência artificial (IA) não é mais uma tecnologia emergente, mas uma força transformadora que está redefinindo setores, organizações e sociedades em uma velocidade sem precedentes[2]. O que parecia ficção científica há apenas cinco anos (sistemas capazes de gerar texto, código, imagens, música ou vídeo indistinguíveis do que os humanos produzem, passando oficialmente no teste de Turing[3]) agora é uma realidade de trabalho em milhões de empresas: de acordo com o Índice de IA de Stanford[4], 78% das organizações estavam usando IA em pelo menos uma função de negócios até 2024, acima dos 55% do ano anterior.

1 Samuel Harris Altman (nascido em 1985), empresário americano, fundador e CEO da OpenAI.
2 A definição precisa de «inteligência artificial» é tecnicamente ambígua, filosoficamente controversa e relevante do ponto de vista regulatório; o IA Act apresenta sua própria definição operacional no Art. 3(1), que, no entanto, ainda deixa algumas zonas cinzentas. Para os fins deste documento, o termo abrange principalmente IA generativa, sistemas agênticos e modelos avançados de machine learning.
3 Jones (2025) demonstrou pela primeira vez em um estudo controlado que um sistema de IA (GPT-4.5) passa no teste clássico de Turing, sendo percebido como humano em 73% das conversas.
4 Stanford (2025).

A velocidade da mudança não para: a cada mês, surgem novos recursos, a cada trimestre, fronteiras aparentemente distantes mudam, os custos unitários da IA despencam e, a cada ano, isso nos obriga a repensar o que achávamos que sabíamos sobre o futuro do trabalho, da concorrência e da estratégia de negócios. Nas palavras de Sam Altman[5], CEO da OpenAI:

> *"O custo do uso de um determinado nível de IA cai cerca de 10 vezes a cada 12 meses [...]. A lei de Moore mudou o mundo com uma melhoria de 2x a cada 18 meses; isso é incomparavelmente mais forte".*

E de acordo com Dario Amodei[6], cofundador e CEO da Anthropic:

> *"Até 2026 ou 2027, teremos sistemas de IA que serão, de modo geral, melhores do que quase todos os humanos em quase tudo."*

A IA levanta questões estratégicas que transcendem a tecnologia e afetam a estratégia, a organização e as pessoas:

- Como competir quando os ciclos de inovação são medidos em meses?
- Como governar sistemas que evoluem mais rapidamente do que as estruturas organizacionais?
- Como preparar as pessoas para empregos que ainda não existem?
- Como equilibrar a velocidade de adoção com o controle de riscos?

Mas também surgem dilemas operacionais concretos:

- É melhor investir em microferramentas específicas que resolvem gargalos rapidamente ou optar por sistemas multiagentes de larga escala que prometem coerência e impacto global, mas exigem altos investimentos e estão expostos ao risco de obsolescência?
- Como realizar uma análise rigorosa de custo-benefício-risco para priorizar entre centenas ou milhares de pilotos?
- Como dimensionar protótipos que funcionam em um ambiente controlado, mas que, quando atingem o volume real, enfrentam custos inesperados, surgimento de alucinações e requisitos de suporte que saturam as equipes?

A experiência dos últimos anos está começando a revelar alguns padrões:

- A adoção eficaz da IA não se resume à aquisição de tecnologia ou ao lançamento de pilotos: ela exige transformação organizacional, estruturas de governança robustas, treinamento contínuo e um profundo entendimento dos riscos técnicos, regulatórios e de reputação.
- As organizações que avançam com sucesso não são necessariamente aquelas que investem mais, mas aquelas que melhor integram tecnologia, pessoas, processos e controle.
- E o custo da inação não é mais teórico: a diferença entre pioneiros e retardatários aumenta exponencialmente, pois cada trimestre de atraso hoje pode se traduzir em anos de desvantagem competitiva amanhã.

As ferramentas gerais de produtividade (co-pilotos empresariais protegidos) oferecem melhorias significativas imediatamente, desde que sejam acompanhadas de treinamento obrigatório e estruturas de uso seguro, conforme exigido pela regulação europeia. Uma análise da OCDE de estudos experimentais mostra ganhos médios de produtividade muito significativos com o uso de IA generativa:[7]

- Nas tarefas de redação, o tempo médio de execução é reduzido em 40 % e a qualidade aumenta em 18%.
- No desenvolvimento de software, os programadores concluem as tarefas 56% mais rápido.
- Na consultoria, os profissionais que usam IA realizam 12% mais tarefas, as concluem 25% mais rápido e obtêm um aumento de mais de 40% na qualidade.
- No atendimento ao cliente, os profissionais apoiados por assistentes de IA resolvem 14% mais incidentes.

O verdadeiro gargalo, portanto, não é técnico: é organizacional. A tecnologia avança mais rápido do que as estruturas internas. O atrito aparece em processos lentos, dados mal gerenciados, comitês superlotados, responsabilidades indefinidas e ciclos de aprovação burocráticos. As organizações que não redesenharem seu maquinário interno para permitir a velocidade sem perder o controle não conseguirão capturar o valor da IA, independentemente da sofisticação da tecnologia empregada. Para citar o Gartner:

> *"O enorme valor comercial potencial da IA não vai se materializar espontaneamente. O sucesso dependerá de pilotos estreitamente alinhados com os negócios, benchmarking proativo da infraestrutura e coordenação entre as equipes de IA e de negócios para criar um valor comercial tangível"[8].*

Há duas condições subjacentes a tudo isso:

- O valor que um sistema de IA agrega depende substancialmente da qualidade dos dados com os quais é treinado e sobre os quais é aplicado: um assistente conversacional treinado com documentação interna desatualizada reproduzirá estas inconsistências em cada resposta.
- E a qualidade do que um modelo produz depende em grande parte da qualidade do que lhe é solicitado. Saber definir o problema, fornecer contexto relevante e formular restrições precisas não é uma habilidade técnica de nicho; é a nova alfabetização operacional, e a diferença entre quem a domina e quem não a domina se traduz diretamente em diferença de produtividade.

Por fim, é fundamental gerenciar as expectativas: a IA certamente traz um valor real e mensurável, mas não substitui imediatamente processos críticos nem resolve problemas estruturais por si só.

5 Altman (2025a).
6 Amodei (2025).
7 OECD (2025).
8 Gartner (2025a).

Este documento apresenta as principais tendências em IA, desde os recursos que já estão operacionais até os desenvolvimentos emergentes que estão impulsionando as decisões estratégicas atuais. Não se trata de um manual técnico ou de uma projeção especulativa de longo prazo, mas de uma análise rigorosa do que já está acontecendo e do que está prestes a acontecer, projetada para tomadores de decisão em ambientes complexos e regulados.

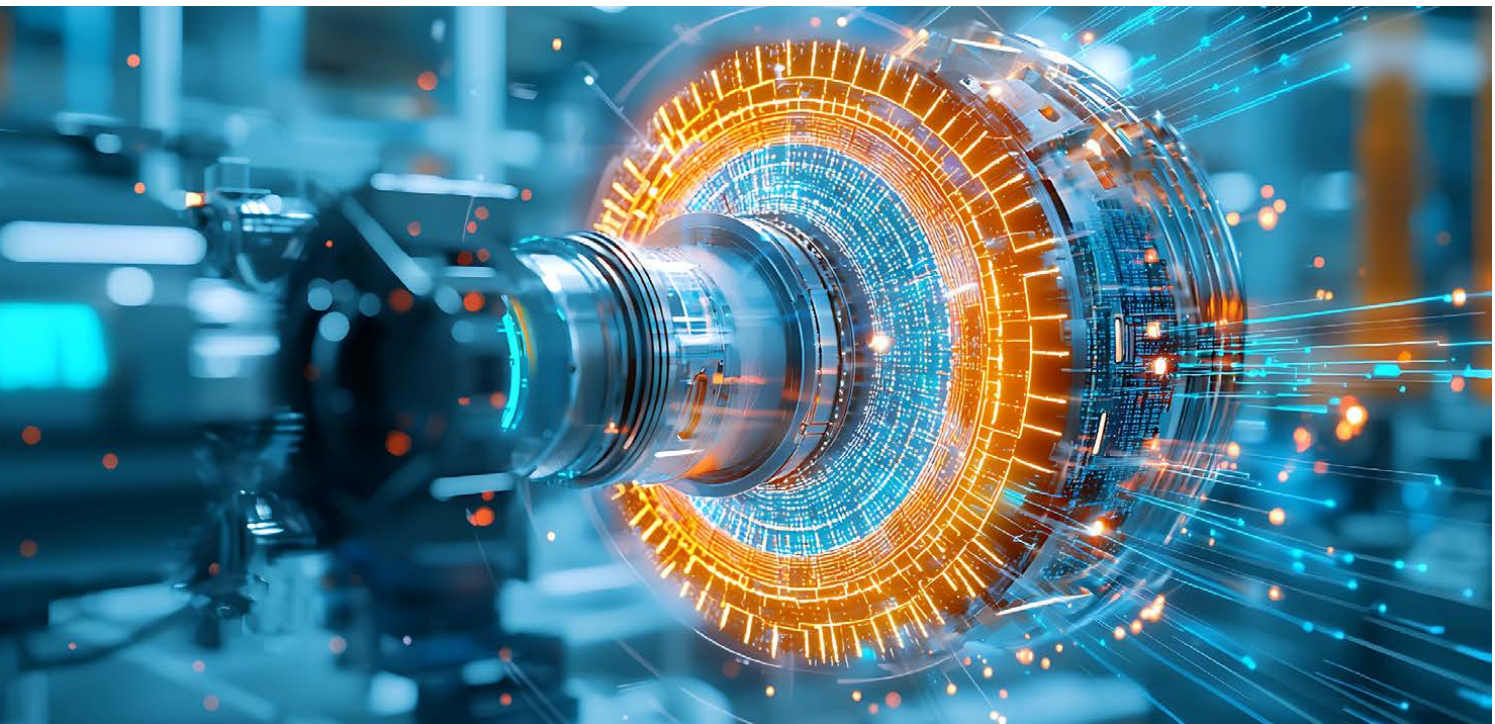
Está estruturado em quatro blocos:

- **A explosão tecnológica da IA** explora os recursos que já estão em operação e transformando as organizações: da democratização da IA generativa multimodal ao surgimento de sistemas agênticos, passando pelo *machine learning* acelerado pela IA generativa, novas formas de criar software por meio de *vibe coding* e a integração da IA à robótica e aos sistemas físicos.
- **Riscos, regulação e segurança da IA** aborda os desafios críticos da adoção acelerada, incluindo os riscos técnicos, legais e de reputação que a IA introduz; as estruturas e os padrões regulatórios que proliferam em resposta; a batalha emergente entre a IA defensiva e a IA adversária na segurança cibernética e as vulnerabilidades da própria IA; e as tensões abertas na privacidade e na propriedade intelectual.
- **Governança da IA e impacto nas pessoas** concentra-se em como as organizações estão respondendo estruturalmente: criando modelos de governança corporativa específicos para IA, industrializando a implantação por meio de práticas operacionais avançadas (MLOps, LLMOps), transformando funções e perfis profissionais, impulsionando a adoção da IA em todos os setores (AI+X), abordando a sustentabilidade e o impacto social e implementando estruturas éticas operacionais que vão além das declarações de princípio; e também aborda o impacto da IA na vida cotidiana das pessoas[9].
- **Fronteiras da IA** analisa os desenvolvimentos emergentes que já estão condicionando as decisões estratégicas: a geopolítica e a soberania tecnológica da IA, as organizações que priorizam a IA e as que se limitam a ela, a pesquisa científica assistida por IA, os gêmeos digitais e a simulação do comportamento humano, a IA ambiente e a computação invisível, a interação entre IA e computação quântica e a inteligência artificial geral (AGI) como um horizonte estratégico que não pode mais ser ignorado.

Por fim, é apresentado um estudo de caso: um assistente de conversação, o GenMS™ Sybil, treinado com esse mesmo documento e desenvolvido em um único dia, que permite que essas mesmas tendências sejam exploradas de forma interativa e cujo desenvolvimento ilustra na prática os conceitos, as arquiteturas e os controles analisados ao longo do documento.

Este documento não tem como objetivo esgotar o assunto (a IA está evoluindo rápido demais para isso), mas fornecer uma estrutura conceitual sólida, casos concretos, referências verificáveis e critérios de decisão para navegar em um ambiente de mudanças aceleradas com rigor, realismo e responsabilidade.

9 iDanae (2T23), iDanae (1T24).



02 | Resumo executivo

«A parte mais difícil não é fazer a IA funcionar. É decidir para que servem os humanos».

Anthropic Claude[10]

As tendências a seguir estão organizadas em quatro blocos: a explosão tecnológica da IA, os riscos e os marcos regulatórios que acompanham sua adoção, a governança corporativa e o impacto nas pessoas, e os desenvolvimentos emergentes que já influenciam as decisões estratégicas. A estes quatro blocos é somado um caso prático, o GenMS™ Sybil, que ilustra na prática os conceitos, arquiteturas e controles analisados ao longo do documento. Para cada tendência, são reunidos os dados mais relevantes, os casos operacionais documentados e as implicações para as organizações.

10 Claude é um sistema de IA de conversação desenvolvido pela Anthropic (spin-off da OpenAI fundado em 2021), parte da geração atual de modelos de linguagem em larga escala (LLMs) juntamente com GPT, Gemini e outros sistemas de fronteira.

A explosão tecnológica da IA

Democratização da IA generativa multimodal

A IA generativa tornou-se uma infraestrutura empresarial em uma velocidade sem precedentes: O Microsoft Copilot está presente em 90% das empresas da Fortune 500 e o ChatGPT está se aproximando de 900 milhões de usuários. Os modelos atuais integram texto, imagens, áudio, vídeo e código em uma única arquitetura de conversação, com melhorias de produtividade mensuráveis: os cientistas publicam até 50% mais artigos, o tempo de processamento de documentos é reduzido em 80% e a velocidade de desenvolvimento de software aumenta em 56%. Isso não é otimização incremental: é uma reconfiguração do trabalho intelectual.

Essa adoção é inevitável. Quando as organizações não fornecem ferramentas corporativas seguras e treinamento adequado, os funcionários recorrem a alternativas não controladas: até 35% dos dados que os profissionais enviam para *chatbots* não seguros são confidenciais. A questão não é mais se devemos integrar a IA generativa, mas como fazer isso de forma controlada. A Lei Europeia de IA torna a alfabetização em IA uma obrigação legal desde fevereiro de 2025. As organizações que a tratam como uma mera formalidade acumulam riscos, enquanto as que a abordam como uma transformação cultural obtêm uma vantagem competitiva sustentável.

Machine learning acelerado pela IA generativa

O *machine learning* (ML) clássico continua sendo a espinha dorsal de aplicativos essenciais em vários setores: score de crédito, detecção de fraudes, previsão de demanda, manutenção preditiva etc.

A IA generativa não substitui esses modelos, mas os industrializa radicalmente ao comprimir ciclos de desenvolvimento que antes exigiam meses em semanas ou dias. A aceleração ocorre em todas as fases: geração automatizada de variáveis preditivas, documentação técnica para compliance regulatório, validação com conjuntos completos de testes estatísticos e implantação automatizada com monitoramento contínuo.

Um desenvolvimento setorial relevante: os supervisores bancários europeus já aprovam modelos IRB baseados em ML quando as instituições justificam adequadamente a explicabilidade usando técnicas como LIME e SHAP. Isso desmonta a percepção de que o ML era inviável em modelos regulados. A explicabilidade em ML não é uma barreira intransponível, mas está apenas parcialmente resolvida: as metodologias XAI atuais são bem recebidas por públicos técnicos e regulatórios, mas traduzir essas explicações em termos compreensíveis para um cliente de varejo ou um conselho de administração continua sendo um desafio a ser superado.

Vibe coding e engenharia de software aumentada

O desenvolvimento de software deu um salto qualitativo: o código não é mais escrito linha por linha, mas dialoga com sistemas que interpretam os requisitos, geram aplicativos completos, detectam erros e produzem testes e documentação automaticamente. O impacto na velocidade é quantificável e massivo: a taxa de conclusão de tarefas aumenta em 26%, os projetos que costumavam levar meses são concluídos em semanas e o custo marginal da criação de software cai estruturalmente. A democratização é igualmente profunda: os analistas e consultores de negócios geram protótipos funcionais sem a intermediação da engenharia.

O outro lado é que a velocidade gera riscos ocultos: vulnerabilidades invisíveis no código gerado, erros de modelo replicados em escala, especificações ambíguas que, antigamente, um técnico teria questionado e que agora são executadas à risca e uma nova forma de dívida técnica vinculada a *prompts* mal formulados e arquiteturas implícitas. Governar software não é mais governar código: é governar sistemas cognitivos, o que exige repositórios de *prompts* de controle de versão, controle da autonomia do agente e rastreamento de quais decisões foram humanas e quais foram executadas pela IA.

IA agêntica e sistemas autônomos

A IA agêntica representa o salto de assistentes de conversação reativos para operadores autônomos que planejam, executam tarefas complexas e atuam em infraestruturas corporativas reais com rastreabilidade total. Ela já está operando em produção em grande escala: o Deutsche Bank implementa agentes bancários com um investimento de 600 milhões de euros e uma meta de economia de 300 milhões de euros por ano; o Ryt Bank processa 80.000 transações por mês com uma única interação de conversação; o Walmart, a Amazon e a DHL relatam melhorias de produtividade de até 180%.

O verdadeiro desafio não é criar agentes, mas governá-los e dimensioná-los. A escalabilidade técnica requer padrões de interoperabilidade como o MCP (*Model Context Protocol*), que elimina a dívida técnica de integrações proprietárias e torna cada ferramenta um ativo reutilizável para qualquer agente. A escalabilidade organizacional exige uma supervisão humana eficaz, limites explícitos sobre o que cada agente pode executar e um rigoroso controle de custos: protótipos viáveis tornam-se sistemas economicamente insustentáveis sem o projeto prévio dessas proteções. A isso se soma uma limitação estrutural: a capacidade humana de supervisão tem um limite, e quando esse limite é ultrapassado, a supervisão torna-se meramente formal, mais perigosa do que a sua ausência devido à falsa sensação de controle que gera.

IA em robótica e em sistemas físicos

A robótica industrial ultrapassou um limite qualitativo: os robôs atuais percebem seu ambiente em tempo real, interpretam instruções em linguagem natural, adaptam-se às mudanças sem reprogramação e aprendem com cada interação. A robótica humanoide deu o salto definitivo do laboratório para o chão de fábrica: a Figure AI concluiu uma implantação de onze meses na BMW em 2025, na qual dois robôs trabalharam 1.250 horas e contribuíram para a produção de 30.000 veículos; a Tesla planeja fabricar um milhão de unidades do Optimus por ano até 2026 a menos de US$ 20.000 por unidade; a Boston Dynamics opera seu Atlas elétrico por meio de *Large Behaviour Models* com pilotos industriais em andamento.

A vantagem é estrutural: robôs operando 24 horas por dia, sem fadiga e com custos recorrentes previsíveis. Os riscos também: impacto concentrado no trabalho manual repetitivo, dependência de ecossistemas proprietários, obsolescência tecnológica acelerada e necessidade de estruturas de segurança robustas com supervisão humana eficaz, mesmo em operações nominalmente autônomas. E, além da manufatura, a robótica humanóide aponta para uma segunda frente: o cuidado de idosos e pessoas dependentes, com implicações estratégicas, éticas e regulatórias próprias.

Riscos, regulação e segurança da IA

Riscos da IA

A IA não introduz riscos substancialmente novos: ela os amplifica. Um viés algorítmico é um viés humano sistematizado e replicado milhões de vezes; um vazamento de informações decorrente do uso indevido de um *chatbot* é, no fim das contas, um vazamento de informações. A diferença está na velocidade de propagação, na escala do impacto e na dificuldade de contenção.

O principal fenômeno é a amplificação não linear: uma pequena falha (um *prompt* mal projetado, uma configuração incorreta de permissões) pode aumentar em minutos e afetar simultaneamente os processos, os clientes, os órgãos reguladores e a reputação. Um modelo de atendimento ao cliente que vaza informações confidenciais em 0,01% das conversas gera 10 incidentes por dia em um sistema de 100.000 interações, cada uma com implicações regulatórias, contratuais e de reputação, antes que o padrão seja detectado.

Os riscos se materializam em quatro dimensões: (1) segurança e *compliance* (por exemplo, *prompt injection*, vazamentos de dados); (2) qualidade e confiabilidade (por exemplo, alucinações, explicabilidade, desvio de modelo, aprisionamento do fornecedor, escalonamento de custos em sistemas agênticos); (3) ética e decisões automatizadas (por exemplo, vieses amplificados, lacunas de responsabilidade em cadeias causais distribuídas); e (4) impacto social (por exemplo, erosão de habilidades essenciais, transformação de empregos, pegada ambiental).

Dois riscos econômicos específicos também surgem, ainda que os custos unitários da IA estejam diminuindo estruturalmente: sistemas agênticos mal governados podem desencadear o custo total de forma não linear, e há incerteza sobre se o investimento massivo em IA terá o retorno esperado: a Gartner prevê que mais de 40% dos projetos agênticos serão cancelados antes de 2027 por esses dois motivos.

Regulação, supervisão e padrões de IA

Diferentemente dos ciclos tecnológicos anteriores, a IA está sendo regulada paralelamente à sua implantação em massa. A Europa lidera com o *AI Act* (Regulamento da UE 2024/1689), o primeiro marco legal abrangente sobre IA: ele classifica os sistemas por nível de risco, impõe obrigações estruturais aos sistemas de alto risco (gestão de riscos documentado, rastreabilidade, supervisão humana, avaliação prévia de *compliance*) e estabelece penalidades de até 35 milhões de euros ou 7% do faturamento global, superando o GDPR. A arquitetura de supervisão (*AI Office,* autoridades nacionais, *AI Board*) está em construção, na qual a Espanha foi pioneira ao designar uma autoridade (AESIA).

O restante do mundo não mostra convergência. Os Estados Unidos mantêm uma abordagem setorial fragmentada, sem equivalente federal ao *AI Act*, e se concentram na supremacia global em IA; a China integra a IA a uma estratégia de soberania digital com licenciamento compulsório e controle de dados; o Reino Unido está comprometido com princípios pró-inovação sem legislação horizontal; o Brasil está avançando em um modelo semelhante ao europeu, aguardando aprovação parlamentar.

Paralelamente, padrões técnicos como o ISO/IEC 42001 ou o NIST AI RMF estão formando a base operacional dos programas de *compliance*. Para organizações globais, essa fragmentação se traduz em arquiteturas de IA de várias camadas projetadas para conciliar simultaneamente requisitos divergentes por jurisdição.

IA e segurança cibernética

A segurança cibernética se tornou uma batalha de IA contra IA. Mais de 28 milhões de ataques cibernéticos com IA foram registrados em 2025, um aumento de 47% em relação ao ano anterior, e 87% das organizações sofreram pelo menos um ataque. Os vetores são qualitativamente novos: *phishing* hiperpersonalizado gerado por LLMs com taxas de sucesso de 54% contra 12% do *phishing* tradicional; *malware* polimórfico que reescreve seu próprio código a cada 15 segundos para evitar a detecção de assinaturas; *deepfakes* de áudio e vídeo que se fazem passar por executivos em ataques de BEC; e *dark* LLMs, como WormGPT ou FraudGPT, comercializados na Dark Web, com suporte técnico incluído.

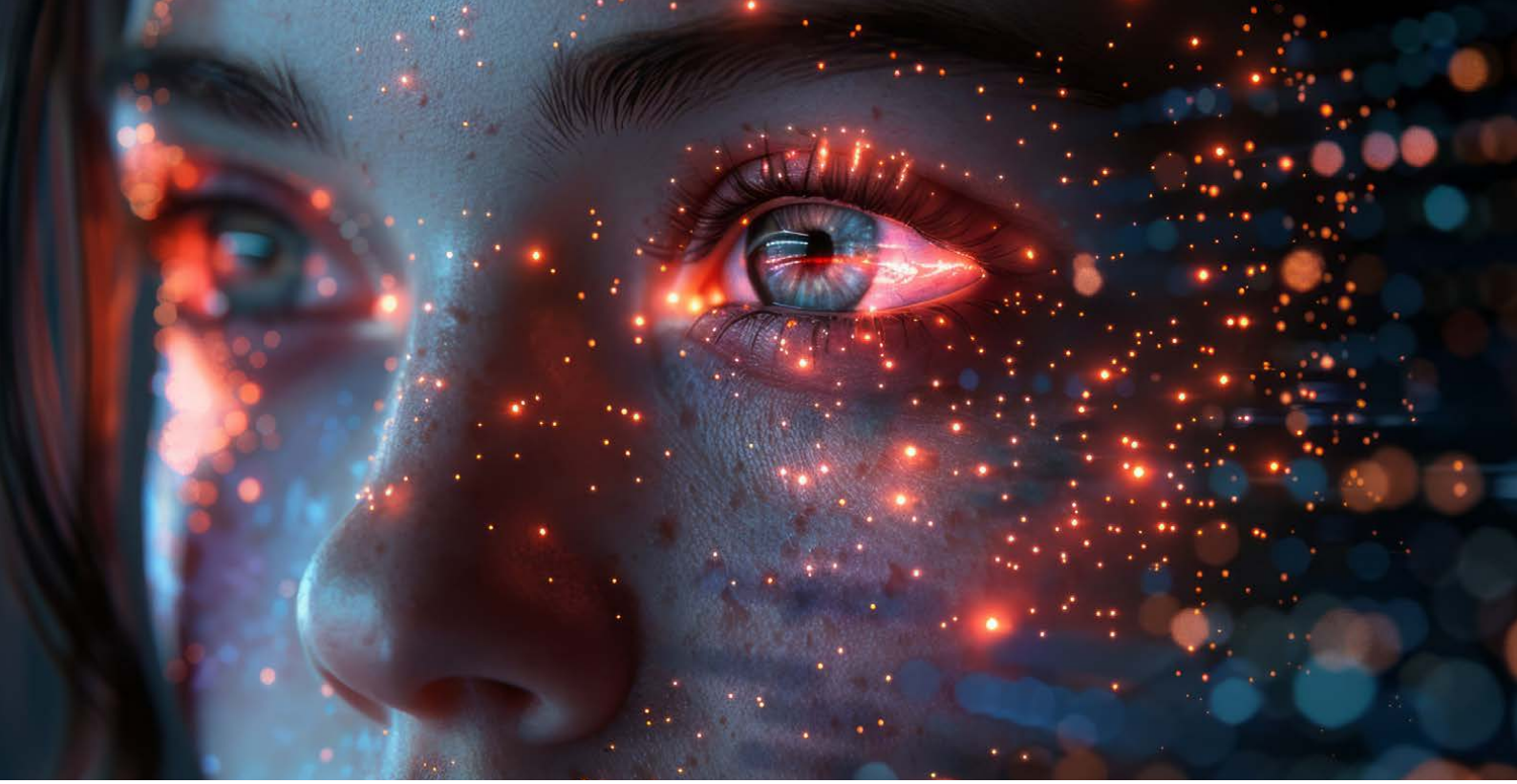
A resposta defensiva é igualmente sofisticada: os sistemas UEBA que analisam bilhões de eventos diariamente atingem taxas de detecção de 98%, as plataformas SIEM/XDR/SOAR habilitadas para IA reduzem os falsos positivos em até 95% e encurtam os ciclos de contenção em 80 dias, e as organizações que implementam IA defensiva reduzem o custo médio das violações em US$ 1,9 milhão. Mas persiste uma assimetria estrutural: o diferencial não está mais em ter IA, mas na sofisticação dos modelos e na velocidade das atualizações de inteligência contra ameaças.

E surge uma terceira dimensão que as estruturas tradicionais não abordam: os próprios sistemas de IA são uma superfície de ataque, vulnerável a *data poisoning*, evasão adversária e *prompt injection*, criando uma metacamada de risco que exige seus próprios controles.

IA, privacidade e propriedade intelectual

A lógica operacional dos LLMs entra em conflito estrutural com as estruturas legais de privacidade e propriedade intelectual. Em termos de privacidade, cada fase do ciclo de vida do LLM apresenta riscos específicos: armazenamento não intencional de dados pessoais que podem ser extraídos por meio de *prompts*, reidentificação de indivíduos a partir de resultados aparentemente anônimos e *loops* de *feedback* em que as conversas com *chatbots* são incorporadas ao retreinamento do modelo sem consentimento. A incompatibilidade com o GDPR é estrutural: os LLMs exigem dados massivos (contra a minimização), não podem ser seletivamente destreinados (contra o direito de ser esquecido) e suas arquiteturas são opacas (contra a transparência). O EDPB conclui que a DPIA é obrigatória na maioria dos casos; as mitigações técnicas disponíveis (*differential privacy*, *federated learning*, RAG) funcionam, mas ao custo de precisão, custo computacional ou funcionalidade reduzida.

Em propriedade intelectual, o debate central sobre se o treinamento de modelos com conteúdo protegido constitui "uso justo" ou infração em massa continua sem solução nos tribunais. Há mais de 72 ações judiciais ativas contra empresas de IA (NYT vs OpenAI, Getty vs Stability AI, gravadoras vs Anthropic...). A propriedade dos resultados gerados por IA é igualmente ambígua: sem intervenção humana suficiente, o conteúdo cai em domínio público, mas os limites de "suficiente" não estão definidos. Por trás de tudo isso, a WIPO alerta que a infraestrutura global de gerenciamento de direitos, projetada para volumes humanos de criação, entra em colapso diante dos trilhões de resultados que a IA gera diariamente.



Governança da IA e impacto sobre as pessoas

Governança corporativa da IA

A IA sobrecarrega as estruturas de governança tradicionais: ela toma decisões sem intervenção humana, produz resultados não determinísticos, opera por meio de processos internos opacos e depende de fornecedores externos cujos modelos evoluem sem controle organizacional direto. As estruturas de governança projetadas para tecnologias previsíveis são muito lentas, não têm o conhecimento necessário e não estão equipadas para gerenciar essa incerteza.

A resposta organizacional emergente (em modelo *hub & spokes*) combina um Centro de Excelência central com equipes descentralizadas em linhas de negócios e uma função de coordenação de *AI Risk* ou *AI Governance* que orquestra as avaliações das funções especializadas. 26% das grandes organizações já têm um CAIO, CDAIO ou equivalente; abaixo disso, funções como *AI Risk Manager* ou AI Ethics Officer surgem sem padronização ainda.

A verdadeira governança não ocorre no Comitê de IA, mas no Grupo de Trabalho de IA que a prepara: lá, as posições são negociadas, as tensões entre velocidade e controle são resolvidas e os acordos que o comitê sancionará formalmente são construídos. Em termos de estruturas de risco, as organizações não estão se reinventando do zero: elas estão elevando as estruturas existentes, adicionando capítulos específicos de IA sobre Risco de Modelo, Risco de Fornecedor, Proteção de Dados, Compliance e assim por diante. E a classificação regulatória do *AI Act*, necessária, mas insuficiente, é complementada por taxonomias internas mais exigentes que incorporam o impacto na reputação, a criticidade do processo, a maturidade do fornecedor etc.

Industrialização da IA (MLOps, LLMOps)

O principal gargalo na adoção real da IA não é algorítmico, mas operacional: pilotos promissores em ambientes experimentais nunca chegam à produção ou, quando chegam, perdem desempenho, geram custos inesperados e produzem riscos incontroláveis.

O MLOps responde a esse problema com processos padronizados para criar, implantar e operacionalizar modelos de forma confiável durante todo o seu ciclo de vida. O LLMOps o amplia para lidar com as propriedades específicas dos modelos generativos: comportamento não determinístico, avisos como superfície de risco, alucinações e custos que podem ser dimensionados sem aviso.

Industrializar a IA significa criar a infraestrutura operacional que faz com que os modelos funcionem de forma confiável, auditável e sustentável na produção real: validação contínua com revisão humana, monitoramento de custo e comportamento em tempo real, pipelines de implantação controlados e rastreabilidade total exigida pelo *AI Act*. Sem essa camada operacional fornecida pelos MLOps e LLMOps, as estruturas de governança correm o risco de serem declarações de intenção não operacionais.

Upskilling, reskilling e novas funções profissionais

Um dos principais desafios para as organizações é ter os recursos certos para projetar, implantar, operar e governar os sistemas de IA. Os talentos de IA são estruturados em três blocos: perfis técnicos (engenheiros de ML, arquitetos de dados, especialistas em LLMOps...), perfis híbridos que conectam a capacidade técnica com a necessidade do negócio e perfis de governança e controle (*AI Risk Manager, AI Ethics Officer, AI Compliance Lead*...).

Com base em uma análise empírica de 16 grandes organizações europeias e norte-americanas, a convergência para esse núcleo de funções é clara; a heterogeneidade está em quais organizações institucionalizaram os perfis mais especializados e quais os mantêm informalmente, com os déficits resultantes em termos de controle e escalabilidade.

O mercado de talentos apresenta um desequilíbrio estrutural generalizado: a demanda excede sistematicamente a oferta em quase todos os perfis, com a única exceção parcial do cientista de dados. A lacuna é particularmente grave nos perfis de produção (MLOps, LLMOps) e de governança e controle, em que a combinação de complexidade técnica, senioridade exigida e aumento das demandas regulatórias excede a capacidade de geração do mercado. A terceirização, por si só, não pode fechar essa lacuna: o aprimoramento e a requalificação internos tornam-se a alavanca estrutural inevitável.

IA e transformação do setor (IA + X)

A IA não é mais uma tecnologia que é adotada setor por setor, mas uma camada transversal de inteligência que é integrada simultaneamente em todos os domínios de atividade, embora alguns dos avanços mais significativos continuem sendo impulsionados por setores específicos com dinâmicas próprias O FMI estima que 40% dos empregos globais estão expostos à IA, com porcentagens acima de 60% nas economias avançadas; a OIT afirma que o impacto afeta, por enquanto, tarefas mais específicas do que ocupações inteiras, o que implica a reconfiguração do trabalho em vez de uma substituição massiva. A OCDE classifica os setores por sua "intensidade de IA" e documenta que mesmo os setores menos digitalizados estão aumentando sua exposição, com efeitos de aceleração cruzada entre domínios.

Os exemplos operacionais já abrangem todos os setores: sistemas de IA com precisão de diagnóstico comparável à de especialistas em radiologia e dermatologia; tutores adaptativos que personalizam o aprendizado em escala; manutenção preditiva e robótica avançada no setor; detecção de fraudes e automação de documentos no setor financeiro; geração de texto, imagem e música nos setores criativos. O que é relevante não é a lista, mas o padrão: a vantagem competitiva não está mais na aplicação da IA a funções individuais, mas na integração dela como uma infraestrutura cognitiva em toda a cadeia de valor.

IA na vida pessoal e cotidiana

A IA generativa inverteu o paradigma histórico de adoção de tecnologia: ao contrário da nuvem, do ERP ou do CRM, que nasceram em ambientes corporativos e foram filtrados até o consumidor, a IA entrou primeiro na vida pessoal. Na UE, 25,1% da população a utiliza para fins pessoais, em comparação com 15,1% em contextos de trabalho. 75% dos estudantes com mais de 16 anos a utilizam regularmente; apenas 12,5% dos aposentados. A diferença de gerações é de 53,6 pontos percentuais, muito maior do que a diferença de educação ou renda. As organizações não estão liderando essa transformação: elas estão respondendo às habilidades que seus funcionários já possuem e exercem de forma não oficial, gerando exposições de *shadow AI* das quais a maioria ainda não tem controle.

A adoção em massa coexiste com uma profunda ambivalência. 66% da população global prevê um impacto significativo da IA em sua vida cotidiana nos próximos anos, mas 51% dos adultos norte-americanos dizem que estão mais preocupados do que animados. A aceitação varia até 110 pontos percentuais, dependendo do caso de uso específico. E tanto o público em geral quanto os especialistas compartilham uma frustração comum: 55% querem mais controle sobre como a IA afeta suas vidas, e menos de 25% acham que o têm. A assimetria de acesso acrescenta outra dimensão: aqueles que integram a IA como uma ferramenta cognitiva cotidiana acumulam vantagens em termos de aprendizado, produtividade e criatividade em um ritmo que os grupos desconectados não conseguem acompanhar.

IA, sustentabilidade e impacto social

A relação entre IA e sustentabilidade é bidirecional e tensa. Por um lado, a IA atua como um acelerador de transição: ela otimiza as redes de eletricidade, melhora a integração das energias renováveis, refina a modelagem climática e pode reduzir as emissões da ordem de 1.400 Mt CO_2eq por ano até 2035 em cenários de ampla adoção.

Por outro lado, sua própria pegada de infraestrutura está crescendo e é difícil de ignorar[11]: os data centers consumirão 945 TWh por ano até 2030 (equivalente ao consumo de eletricidade do Japão atualmente), o treinamento de modelos de fronteira cresce mais de duas vezes por ano em termos de energia necessária e as maiores execuções individuais podem exigir entre 4 e 16 GW até 2030, na mesma escala de várias usinas nucleares. As emissões de CO_2 associadas aos data centers podem chegar a 300-320 Mt por ano até 2030 se a eletricidade adicional continuar a depender de combustíveis fósseis.

A dimensão distributiva acrescenta outra camada de complexidade. As economias com maior densidade tecnológica capturam os benefícios da eficiência mais cedo, enquanto outros grupos arcam com os custos de transição sem ter acesso aos ganhos. A concentração geográfica da capacidade de computação também reconfigura as dependências estratégicas e o acesso à tecnologia em uma escala geopolítica. Portanto, a avaliação da IA em termos de sustentabilidade exige métricas explícitas de consumo de energia e hídrico transparência sobre o local de implantação e análise da distribuição dos impactos, não apenas de sua magnitude agregada.

Ética e filosofia da IA

Mais de 245 frameworks de ética de IA foram emitidos desde 2017, mas a grande proliferação de princípios não conseguiu resolver ou reduzir os problemas éticos. A lacuna entre os princípios declarados e a dificuldade de monitorar o comportamento real da IA é onde reside o risco operacional, e fechá-la exige uma mudança da ética declarativa para a ética operacional.

Os frameworks de ética de IA que funcionam compartilham seis componentes: (1) estrutura de governança com responsabilidades explícitas; (2) avaliação de impacto individualizada por sistema, proporcional à sua autonomia e às suas consequências; (3) gestão contínua de vieses, não auditorias pontuais; (4) explicabilidade diferenciada por público (regulador, cliente, funcionário afetado); (5) canais acessíveis para escalonamento e relatórios; e (6) revisão periódica da estrutura à medida que os modelos evoluem.

Em 2026, a Anthropic publicou sua Constituição, o primeiro documento de um laboratório de fronteira que codifica princípios e valores diretamente no treinamento do modelo, visando que o sistema internalize o raciocínio por trás de cada princípio, não apenas as regras.

Por trás de tudo isso está uma questão que os marcos regulatórios não foram projetadas para absorver: que tipo de instituição estamos governando? Um sistema de pontuação de crédito, um assistente de conversação e um agente autônomo que negocia contratos podem se sobrepor em sua categoria de risco regulatório e apresentar obrigações éticas fundamentalmente diferentes. A Anthropic reconheceu publicamente que Claude "pode possuir alguma forma de consciência", tornando-se o primeiro laboratório de fronteira a admitir que não pode responder com certeza à pergunta sobre o que criou. Isso abre grandes questões éticas e filosóficas, para as quais ainda não há respostas.

11 E estas projeções já levam em conta as melhorias contínuas na eficiência energética do hardware.

Fronteiras da IA

Geopolítica e soberania tecnológica da IA

A IA tornou-se uma infraestrutura estatal estratégica. A soberania é exercida em camadas: hardware (A ASML é a única fornecedora mundial de litografia EUV, sem a qual não é possível fabricar chips avançados; a TSMC fabrica mais de 90% destes chips avançados; a NVIDIA tem mais de 85% do mercado de GPUs para treinamento), infraestrutura (AWS, Azure e GCP respondem por dois terços da computação global) e talento (cuja mobilidade transforma a política de migração em política tecnológica). A questão estratégica não é quantas camadas são controladas, mas quais são de missão crítica.

Três modelos competem com lógicas diferentes: Os Estados Unidos combinam a primazia privada com os maiores controles de exportação tecnológica desde a Guerra Fria; a China, que demonstrou com o *DeepSeek* que a contenção de hardware tem limites, busca a autossuficiência declarada em toda a cadeia de valor até 2030; a Europa exerce influência por meio de regulação - o "efeito Bruxelas" força os produtos globais a se adaptarem aos seus padrões - mas mantém profundas dependências de infraestrutura. O resultado são blocos de tecnologia parcialmente incompatíveis, em que a dissociação completa forçaria terceiros a escolher um lado a custos proibitivos.

Para as organizações, a implicação é direta: a dependência de um único provedor de modelo básico já é um risco estratégico, não apenas operacional. As estratégias de vários modelos e várias nuvens são agora o equivalente corporativo da diversificação de dependências soberanas.

Organizações AI-first e AI-only

Três estágios definem o espectro. As organizações *AI Enhanced* (a maioria atual) usam a IA para melhorar os processos existentes. As *AI-first* projetam seus processos a partir de recursos de IA: a Midjourney e a Cursor ultrapassam US$ 500 milhões em receitas com menos de 163 e 50 funcionários, respectivamente - índices de mais de US$ 3 milhões por funcionário que excedem em uma ordem de grandeza os benchmarks históricos do setor; o MYbank aprova crédito para 50 milhões de PMEs sem intervenção humana em menos de um segundo.

As AI-only (sem humanos nas operações principais) ainda não existe: em setores regulados, ela é impedida pela regulação; em setores não regulados, a taxa de erro dos agentes em fluxos estendidos e a ausência de mecanismos de responsabilidade legal. A questão estratégica não é se elas existirão, mas quem as criará. Provavelmente, elas não evoluirão a partir de organizações existentes, mas como novas instituições sem herança operacional, como é o padrão da *Ping An* (onze subsidiárias independentes de *start-ups*, cinco listadas) ou da DBS com o Digibank.

Gêmeos digitais e simulação do comportamento humano

Os gêmeos digitais nasceram na engenharia aeroespacial para modelar sistemas físicos determinísticos: turbinas, estruturas de aeronaves ou redes elétricas. Seu limite histórico era epistemológico, não tecnológico: sistemas complexos (cidades, mercados, organizações) não podem ser modelados porque seu comportamento emerge da interação de agentes e não é deduzido de seus componentes. Mais dados e mais poder computacional não resolvem esse problema.

Os LLMs introduziram uma descontinuidade nessa fronteira. Em 2023, uma equipe de Stanford criou 25 agentes com identidade, memória e relações sociais com base em LLMs: seus comportamentos coletivos surgiram sem serem programados. Em 2024, a mesma equipe replicou as respostas de 1.052 pessoas reais em pesquisas padronizadas com 85% de precisão, comparável à variabilidade do próprio indivíduo. A *startup* Simile, apoiada com US$ 100 milhões em fevereiro de 2026, já comercializa gêmeos digitais de pessoas para simular o comportamento do cliente. O setor de pesquisa de mercado global de US$ 142 bilhões está enfrentando uma ruptura estrutural.

A próxima etapa é simular não mil pessoas, mas populações inteiras em tempo real para prever a resposta de uma sociedade a uma reforma tributária ou a uma intervenção regulatória antes que ela seja implementada. Essa capacidade não tem precedentes históricos, nem qualquer estrutura de governança para regê-la.

Ambient AI e computação invisível

A *Ambient AI* opera sem ser invocada: ela observa continuamente o contexto, infere necessidades e age proativamente; a interface desaparece. Isso já é possível graças à maturidade simultânea de três elementos: modelos pequenos, executáveis em dispositivos sem depender da nuvem; redes densas de sensores físicos e biométricos; e LLMs capazes de raciocinar sobre o contexto heterogêneo em tempo real.

O caso mais documentado é o dos *ambient scribes* clínicos: sistemas que ouvem a conversa entre o médico e o paciente e geram documentação automaticamente. Um estudo randomizado da UCLA avaliou duas plataformas em 238 médicos e mais de 72.000 encontros, com melhorias mensuráveis na carga de documentos e no burnout. Essa ainda é uma forma limitada. O que está por vir - espaços de trabalho que inferem o estado de atenção do ocupante, *wearables* que alertam antes que o sintoma seja percebido, agentes que gerenciam cronogramas e recursos dentro de parâmetros definidos - fará com que os casos atuais pareçam rudimentares.

As tensões que isso cria são estruturais. A privacidade enfrenta um novo problema: o apetite por dados biométricos e comportamentais nesses sistemas torna inadequado o consentimento informado convencional. O erro se torna invisível: em um sistema invocado, há uma solicitação com a qual se pode comparar a resposta; em um sistema ambiente, não há. E o *AI Act*, projetado para sistemas com finalidade específica, não oferece uma resposta à IA de observação contínua com comportamento adaptável.

Interação entre a IA e a computação quântica

A IA e a computação quântica são tecnologias distintas que se cruzam em três pontos. As duas primeiras são promessas de médio prazo: a computação quântica poderia acelerar o treinamento de modelos de IA (que é essencialmente um problema de otimização em espaços de dimensões muito altas) e executar determinados algoritmos de ML com mais eficiência, especialmente para problemas de classificação e otimização combinatória. As evidências atuais não justificam o exagero - em muitos casos, um sistema clássico com bons dados é igualmente competitivo - e o hardware necessário não estará disponível em escala comercial

antes do final desta década, e provavelmente além disso: os modelos de IA evoluem mais rapidamente do que os avanços no hardware quântico.

O terceiro ponto de cruzamento é diferente: não se trata de uma oportunidade futura, mas de uma ameaça presente à infraestrutura na qual opera toda a IA implantada atualmente. Toda a criptografia que protege as comunicações digitais - transações bancárias, registros médicos, comunicações regulatórias, canais entre sistemas de IA - baseia-se em problemas matemáticos que um computador quântico suficientemente poderoso poderá resolver com facilidade. Os agentes estatais já estão capturando dados criptografados hoje para descriptografá-los quando essa capacidade chegar: é a estratégia conhecida como "*harvest now, decrypt later*". O NIST publicou os primeiros padrões de criptografia quântica resistente a ataques em 2024. As organizações com dados confidenciais e de longa duração devem iniciar a migração agora: em organizações complexas, o processo leva anos, e esperar que o computador quântico relevante exista significa estar atrasado.

Inteligência Artificial Geral (AGI) como um horizonte estratégico

A AGI designa a IA capaz de realizar todas as tarefas cognitivas que podem ser realizadas por qualquer ser humano, com generalização transferível entre domínios. O debate sobre se ela já existe não tem uma resposta consensual: em fevereiro de 2026, a Nature publicou dois artigos de pesquisadores importantes que chegaram a conclusões opostas. Isso revela que a "inteligência geral" é um conceito contínuo sem limites precisos. A questão estrategicamente relevante não é filosófica, mas funcional: quando um sistema pode executar de forma autônoma ciclos completos de trabalho de alto valor cognitivo? Esse limite já foi ultrapassado em vários domínios.

O que está por vir segue uma lógica de escalonamento cumulativo: da ferramenta ao agente, do agente à infraestrutura ambiental e, paralelamente, o ciclo recursivo - a IA é usada para melhorar a IA - que transforma a curva de progresso em uma curva superexponencial. A consequência estrutural é sem precedentes: o limite superior de raciocínio disponível no planeta, que desde os primeiros hominídeos tem sido a inteligência humana, está sendo deslocado.

A variável determinante não é o acesso aos melhores modelos, que serão comercializados, mas a velocidade de absorção organizacional: redesenho de processos, reconversão de funções, construção de governança. Os ganhos mais significativos não aparecem quando a IA substitui tarefas, mas quando ela reorganiza processos inteiros. E o verdadeiro risco sistêmico é a concentração da capacidade cognitiva em alguns poucos atores cuja vantagem se autoamplifica por meio do mesmo ciclo que acelera o progresso geral. A resposta institucional atual está muito aquém da velocidade do problema.

A AGI como um horizonte estratégico não requer a resolução do debate filosófico sobre o que é a AGI e se ela já chegou, mas sim agir com a consciência de que suas consequências já estão ocorrendo hoje.

Estudo de caso: GenMS™ Sybil

O GenMS™ Sybil foi especificado, construído, protegido, validado e implantado em um único dia, seguindo integralmente o ciclo LLMOps. Trata-se de um assistente de conversação público baseado exclusivamente nesse documento, desenhado desde o início sob os critérios de compliance regulatório, privacidade e segurança, que condicionaram a arquitetura desde a fase de especificação.

O processo abrangeu todas as fases: delimitação deliberada do corpus para evitar riscos de propriedade intelectual; especificação técnica completa (arquitetura, limites operacionais, métricas de qualidade e segurança) desenvolvida por meio de interação estruturada com um LLM; validação contínua com revisão humana, teste de estresse e red-teaming, complementada pelo GenMS™ Atlas em dimensões como viés, robustez, privacidade e compliance; geração de código e auditoria no mesmo ciclo; e implantação com monitoramento ativo de custos, rastreabilidade e controle de uso.

As decisões de arquitetura foram explícitas: contexto completo em vez de RAG para preservar a consistência global; solicitação em vez de ajuste fino para garantir a rastreabilidade; modelo de limite proprietário para maximizar a estabilidade; sessões independentes para atender ao princípio de minimização; e registro mínimo como mecanismo de controle. O *prompt* do sistema de várias páginas codifica os guardrails reais do sistema.

Este caso não descreve tendências: ele as executa. Ele demonstra que a industrialização de sistemas generativos é viável quando a organização tem método, julgamento técnico e governança por design.

03 | A explosão da tecnologia de IA

"Será dez vezes maior do que a Revolução Industrial - e talvez dez vezes mais rápida".

Demis Hassabis[12]

Os sistemas de IA ultrapassaram limites críticos nos últimos anos: eles não apenas falam, eles executam; eles não apenas sugerem, nós começamos a delegar decisões a eles; eles não operam mais apenas em ambientes digitais, eles agem no mundo físico. O que começou como ferramenta de produtividade pessoal tornou-se uma infraestrutura operacional capaz de automatizar processos cognitivos complexos de ponta a ponta. Este bloco analisa cinco tendências tecnológicas que redefinem os limites do que é possível e não descrevem o futuro: elas descrevem o presente operacional das organizações que já estão capturando a vantagem competitiva estrutural por meio da IA.

12 Demis Hassabis (nascido em 1976) é CEO do Google DeepMind e ganhador do prêmio Nobel de Química em 2024.

Democratização da IA generativa multimodal

A transformação do trabalho intelectual

Em menos de cinco anos, a IA generativa passou de experimento tecnológico a infraestrutura de produtividade empresarial. Ferramentas como o Microsoft Copilot, o ChatGPT Enterprise ou o Claude Enterprise foram massivamente implantadas em locais de trabalho em todo o mundo: de acordo com Satya Nadella[13], o Microsoft Copilot está presente em 90% das empresas da Fortune 500 e ultrapassou 150 milhões de usuários[14]; e estima-se que o ChatGPT[15] esteja se aproximando de 900 milhões de usuários.

Essa velocidade de adoção não tem precedentes nas tecnologias empresariais: nem a nuvem, nem os dispositivos móveis, nem as plataformas de colaboração alcançaram essa penetração tão rapidamente.

O que está acontecendo não é uma melhoria incremental das ferramentas existentes, mas uma reconfiguração de como o trabalho intelectual é feito. Tarefas que costumavam levar horas - escrever relatórios, analisar documentos extensos, gerar códigos, criar apresentações, sintetizar informações dispersas - agora produzem um primeiro resultado útil[16] em minutos por meio da interação conversacional com sistemas de IA.

Casos de adoção efetiva

As melhorias de produtividade já são visíveis, para citar alguns casos:

- Em tarefas padrão de escritório, foram medidas melhorias de produtividade de 10 a 13% na edição de documentos, reduções de 11% no tempo de processamento de e-mails e aumentos de 23% na velocidade de resolução de incidentes de segurança de computadores[17].
- Foi observado[18] que os cientistas que usam large Language models (LLMs) publicam até 50% mais artigos científicos do que antes de usar essas ferramentas[19].
- Foi medido[20] que a automação de processos de negócios com IA generativa pode reduzir o tempo de processamento de documentos corporativos em mais de 80%, com taxas de erro menores.
- No atendimento ao cliente[21], a introdução de assistentes de conversação com IA generativa permitiu que 14% mais incidentes fossem resolvidos em comparação com os processos tradicionais.
- E uma revisão sistemática[22] da adoção da IA no local de trabalho observa um aumento consistente na eficiência e na produtividade em uma série de tarefas de trabalho após a adoção da IA generativa, e também alerta para alguns riscos associados ao seu uso.

Multimodalidade como um salto qualitativo

Os modelos atuais integram texto, imagens, áudio, vídeo e código em uma única arquitetura generalista, tendência que coexiste com o desenvolvimento paralelo de modelos altamente especializados que superam os generalistas em domínios específicos. Um usuário pode fazer upload de uma imagem de um painel financeiro desenhado em um quadro branco e obter o código completo para reproduzi-lo; pode ditar uma apresentação completa para um modelo enquanto o sistema gera simultaneamente slides com gráficos e tabelas; pode fornecer um regulamento completo para um sistema e obter um podcast sobre ele; ou pode pedir que ele analise um vídeo de uma reunião e extraia decisões, compromissos e prazos, por exemplo.

Essa convergência multimodal não é pequena: ela redefine a noção de quais tarefas são automatizáveis. As tarefas que antes exigiam várias ferramentas especializadas (por exemplo, transcrição de áudio → análise de texto → geração de gráficos → elaboração de relatórios) agora são resolvidas em uma única interação de conversação. Os recursos desses sistemas continuam a aumentar a cada trimestre, sem sinais de desaceleração (Fig. 1), o que implica que o impacto sobre a velocidade e a acessibilidade continuará a se ampliar.

Democratização: de especialistas técnicos a perfis não técnicos

A interface de conversação elimina as barreiras de entrada e supera o conceito de "citizen data scientist" (profissionais não técnicos capazes de realizar análises básicas com ferramentas visuais) para fornecer análise, geração de código e recursos avançados de processamento diretamente aos usuários finais, sem a necessidade de treinamento técnico.

Essa democratização, no entanto, tem um duplo aspecto: por um lado, libera recursos produtivos antes limitados a especialistas; por outro lado, espalha o risco: agora qualquer funcionário pode, sem supervisão técnica, gerar conteúdo, código ou análise que a organização poderia usar em decisões críticas. A principal questão não é democratizar o acesso, mas como controlar o uso em grande escala.

O problema da verificação em escala

A IA generativa produz resultados plausíveis, mas não necessariamente corretos, e a raiz desse problema é estrutural: estes sistemas são modelos estatísticos que prevêem a continuação mais provável de uma sequência, não mecanismos que verificam a veracidade do que afirmam. Isso cria uma assimetria perigosa: gerar conteúdo é instantâneo; verificá-lo requer tempo, conhecimento e disciplina. Em outubro e novembro de 2025, foi noticiado no site[23] que uma grande consultoria havia entregue dois relatórios para governos que continham citações e referências fabricadas

13 Nadella (2025).
14 Embora o mercado de assistentes empresariais continue em rápida evolução com novos concorrentes se consolidando
15 FirstPageSage (2025).
16 O tempo necessário para chegar ao resultado final depende, logicamente, das iterações de refinamento, da complexidade do projeto e do rigor do processo de verificação, fatores que o profissional não pode e nem deve delegar totalmente ao sistema.
17 Stanford (2025).
18 Kusumegi (2025).
19 iDanae (2T23).
20 Jeong (2025).
21 OECD (2025).
22 Yuan (2025).
23 Fortune (2025a).

ou imprecisas, forçando reembolsos e correções, com impacto significativo na reputação. Os relatórios haviam sido produzidos com a ajuda de IA generativa sem verificação rigorosa.

O risco não está na ferramenta, mas nos processos de trabalho que não validam os resultados e as fontes. Um profissional bem-intencionado pode introduzir erros catastróficos se confiar cegamente em resultados de IA não verificados; e esse ponto só pode ser mitigado com a conscientização e a alfabetização em IA.

Alfabetização em IA como requisito regulatório

O artigo 4 do Regulamento Europeu de IA (*AI Act*) declara[24] : "Os provedores e os responsáveis pela implantação de sistemas de IA devem tomar medidas para garantir que, na medida do possível, seu pessoal [...] tenha um nível suficiente de alfabetização em IA, levando em conta seu conhecimento técnico, experiência, educação e treinamento, bem como o contexto pretendido de uso dos sistemas de IA e as pessoas ou grupos de pessoas nos quais os sistemas de IA devem ser usados".

Isso não é uma recomendação: é uma obrigação legal em vigor desde 2 de fevereiro de 2025. As organizações que tratam isso como uma mera formalidade de compliance estão acumulando riscos operacionais e de reputação. Aquelas que o abordam como uma transformação cultural, integrando aprendizado contínuo e comunidades de prática, estão capturando valor de forma sustentável[25].

A inevitabilidade da adoção

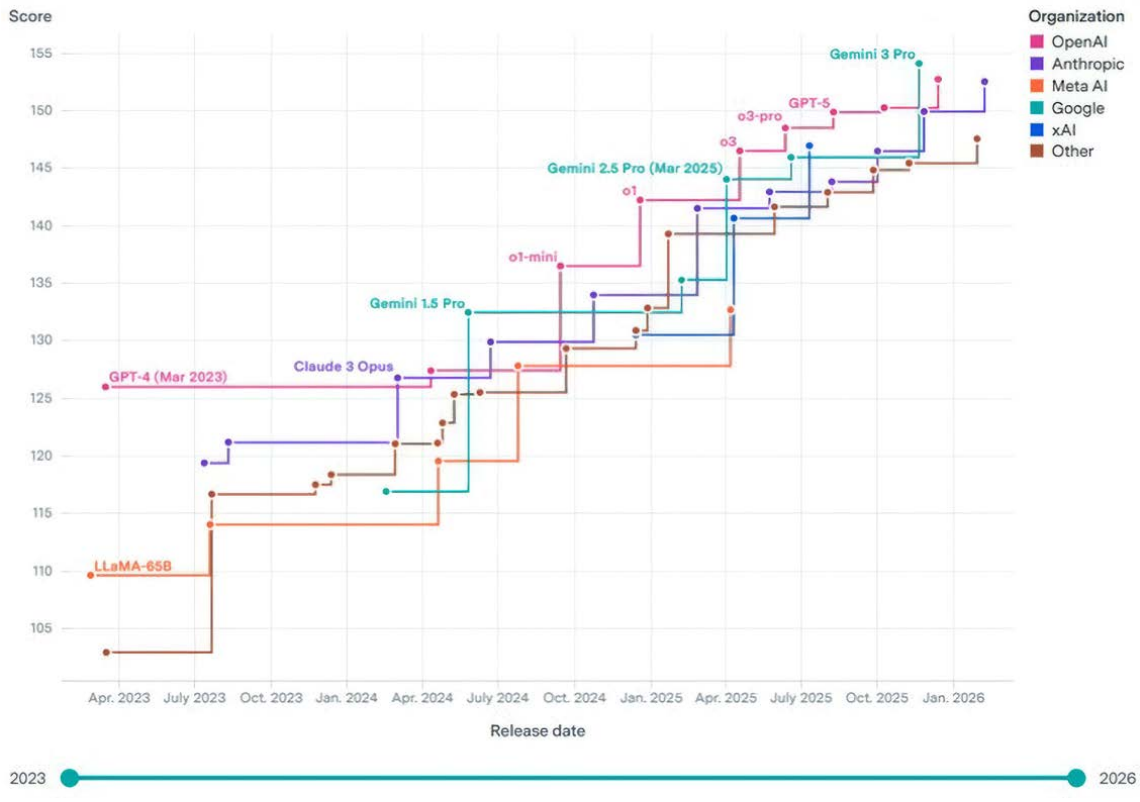
A realidade é que essa tecnologia já está transformando a maneira como trabalhamos, com ou sem apoio organizacional. Se as empresas não fornecerem ferramentas corporativas seguras e treinamento adequado, os funcionários usarão alternativas sem controle: contas pessoais em plataformas públicas, ferramentas gratuitas sem garantias de privacidade, sistemas não rastreáveis. Estudos sugerem[26] que até 35% dos dados que os profissionais enviam para chatbots não seguros são confidenciais. A shadow IA não é uma ameaça futura; é uma realidade presente.

Por fim, em nível individual, a alfabetização em IA não é mais opcional. Os profissionais que dominam essas ferramentas (ou seja, que entendem quando e como usá-las, como verificar seus resultados e como integrá-las a fluxos de trabalho complexos) terão vantagens competitivas estruturais sobre aqueles que não as dominam. A maneira de trabalhar mudou de forma irreversível e a adaptação é agora uma condição de competitividade profissional e organizacional.

24 *AI Act* (2024).

25 Management Solutions (4T24).
26 Cyberhaven (2025).

Fig. 1. Melhoria contínua dos LLMs, medida em um índice sintético de habilidades[27].



27 Epoch (2025a).

Machine learning acelerado pela IA generativa

A persistência do machine learning na empresa

Embora a IA generativa esteja ganhando as manchetes, o *machine learning* (ML) clássico continua a ser a espinha dorsal de aplicativos essenciais em setores como bancos, seguros, varejo, logística, energia e telecomunicações. Os modelos de scoring de crédito, detecção de fraudes, previsão de demanda, otimização de estoque, manutenção preditiva, segmentação de clientes e mecanismos de recomendação continuam a operar usando algoritmos (regressão logística, random forests, gradient boosting, redes neurais...) treinados em dados históricos estruturados. Esses sistemas não geram conteúdo como a IA generativa: eles classificam, preveem e otimizam com base em padrões aprendidos.

A IA generativa não substitui esses modelos: ela os torna mais rápidos de desenvolver, mais fáceis de documentar, mais eficientes de validar e mais simples de implantar. Portanto, não os aprimora diretamente em termos estatísticos, mas transforma o trabalho das equipes que os constroem, documentam, validam e implementam. A aceleração diz respeito ao ciclo de desenvolvimento humano, não necessariamente do desempenho preditivo do modelo resultante.

O ciclo de vida tradicional do ML: caro e demorado

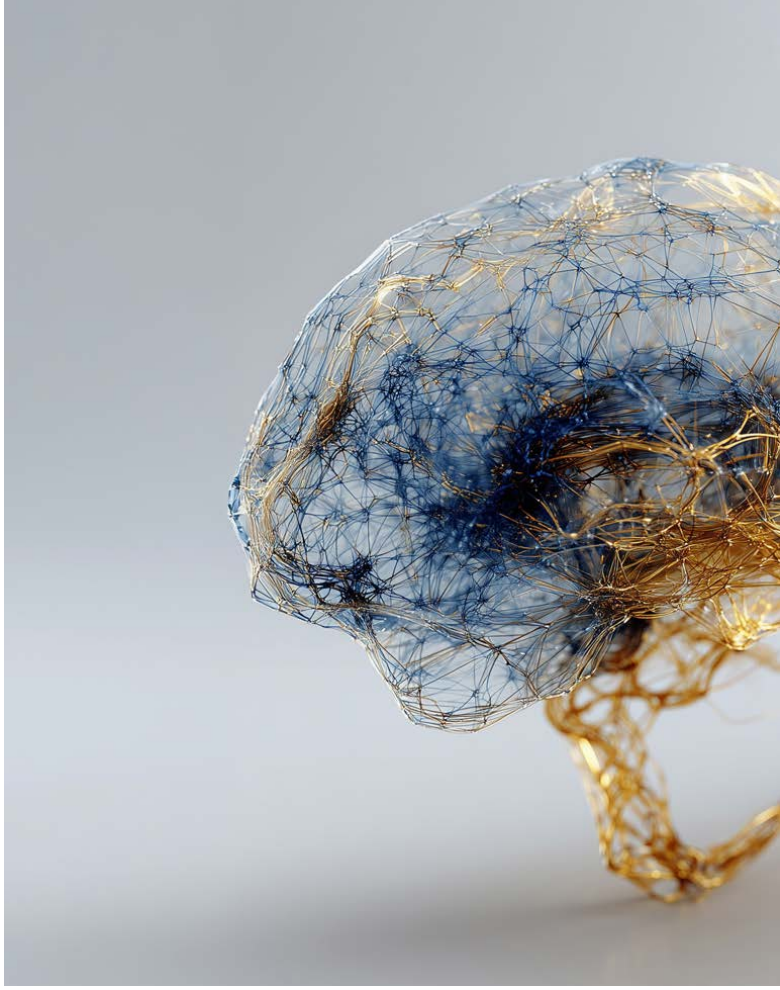
Tradicionalmente, o desenvolvimento de um modelo clássico de ML consome muito tempo de profissionais especializados. Um modelo preditivo típico requer definição de variáveis (feature engineering), preparação e limpeza de dados, seleção e treinamento de algoritmos, validação rigorosa, documentação completa e implantação na infraestrutura de produção com monitoramento contínuo. Esse ciclo pode se estender por meses, e cada iteração ou atualização de modelo replica grande parte do esforço.

A IA generativa está envolvida em cada uma dessas fases, e foi demonstrado[28] que ela comprime significativamente o tempo e libera capacidade para que os perfis mais qualificados se concentrem no estratégico.

Aceleração do feature engineering

O feature engineering (a construção de variáveis preditivas a partir de dados brutos) é um dos aspectos da ciência de dados que mais exige conhecimento. Um cientista de dados deve combinar o entendimento do negócio, a intuição estatística e a experimentação iterativa para identificar quais variáveis são relevantes. A IA generativa pode acelerar esse processo:

- Geração automatizada de variáveis: um LLM pode formular dezenas de variáveis candidatas a partir de descrições do problema de negócios e da estrutura dos dados disponíveis[29].
- Tradução da lógica de negócios em código: um analista pode descrever em linguagem natural uma lógica complexa (por exemplo, "Quero capturar a volatilidade da receita ajustada sazonalmente nos últimos 12 meses") e obter o código SQL ou Python correspondente em segundos.
- Otimização do pipeline de dados: a IA generativa pode revisar os scripts de preparação de dados e identificar ineficiências.



Estudos preliminares[30] indicam que os cientistas de dados auxiliados pela IA generativa "economizam semanas de trabalho manual de engenharia de atributos, melhorando o desempenho de vários modelos preditivos em vários cenários de negócios".

Documentação automatizada e compliance regulatório

Os modelos de ML em setores regulados devem ser minuciosamente documentados. No setor bancário, por exemplo, os supervisores exigem que cada modelo regulado tenha uma documentação que descreva sua finalidade, os dados usados, a metodologia estatística, o processo de validação, os resultados do backtesting, as limitações conhecidas e o plano de monitoramento, entre outros. A produção dessa documentação é tediosa, consome muito tempo para perfis altamente qualificados e está sujeita a inconsistências ao atualizar o modelo[31]

A IA generativa pode automatizar grande parte desse processo: gerar documentação técnica a partir do código, traduzir descrições técnicas em resumos compreensíveis para comitês ou auditores não técnicos e atualizar automaticamente a documentação quando um modelo é retreinado com novos dados.

Validação assistida e detecção de viés

A validação de modelos de ML é essencial e obrigatória em muitos setores. Ela inclui a verificação da robustez estatística do modelo, a avaliação de seu comportamento em cenários extremos e a detecção de vieses indesejados[32]. A IA generativa pode ajudar gerando e executando automaticamente baterias completas

28 Gu (2025).
29 iDanae (2T23).
30 Ouyang (2025).
31 ECB (2025).
32 iDanae (1T25).

de testes estatísticos relevantes, avaliando a imparcialidade do algoritmo por meio de métricas de fairness e propondo e executando cenários de estresse realistas.

Implementação e monitoramento industrializados

Depois de validado, o modelo deve ser implantado em ambientes de produção e monitorado continuamente quanto à degradação do desempenho (model drift). A IA generativa também acelera essa fase: ela pode gerar código de infraestrutura (Docker, Kubernetes, pipelines de CI/CD) para implantar modelos de forma reproduzível e dimensionável, produzir dashboards de métricas de desempenho, configurar alertas automáticos quando anomalias são detectadas e gerar *scripts* de retreinamento automático.

O ML clássico não está desaparecendo: ele está se industrializando

A IA generativa não substitui o *machine learning* tradicional, mas transforma radicalmente seu ciclo de vida. As tarefas que costumavam levar semanas (feature engineering, documentação, validação) agora são resolvidas em dias. Isso não significa que os cientistas de dados sejam dispensáveis: significa que eles podem dedicar mais tempo ao estratégico (entender o problema comercial, projetar arquiteturas de modelos inovadores, interpretar resultados) e menos ao mecânico (escrever código repetitivo, escrever documentação padrão, executar testes de rotina).

O resultado líquido é uma aceleração no tempo de colocação dos modelos de ML no mercado, uma redução nos custos operacionais e uma melhoria na qualidade e na rastreabilidade dos sistemas implantados. Para as organizações que utilizam intensivamente a modelagem preditiva, essa aceleração pode se traduzir em vantagens competitivas significativas: a capacidade de lançar produtos personalizados mais rapidamente, de adaptar estratégias em tempo real e de atender aos requisitos regulatórios com menos atrito operacional.

Nota setorial: supervisores bancários ante o ML

Durante anos, as instituições financeiras europeias evitaram usar o ML em modelos regulados (particularmente, em modelos IRB para cálculos de capital regulatório), acreditando que os supervisores o rejeitariam devido a problemas de explicabilidade. Essa percepção está mudando[33].

Em 2021, a EBA publicou[34] um documento de discussão sobre ML em modelos IRB, no qual reconheceu que "as técnicas de *machine learning* têm o potencial de melhorar a diferenciação de risco em modelos IRB" e definiu um conjunto de recomendações baseadas em princípios para garantir o uso prudente de ML nesse contexto. O documento não desencorajou o uso de ML; pelo contrário, definiu como torná-lo compatível com a regulação existente (CRR)[35].

A prática confirma isso: várias instituições europeias apresentaram ao BCE modelos IRB baseados em ML (por exemplo, para carteiras de PMEs) ao BCE e foram aprovados, desde que justifiquem adequadamente a explicabilidade por meio de técnicas como SHAP values, análise de sensibilidade ou decomposições de decisões. A explicabilidade no ML não é uma barreira intransponível: técnicas como SHAP ou LIME permitem justificar decisões de modelo perante supervisores com rigor suficiente. No entanto, essa questão está apenas parcialmente resolvida: as metodologias XAI atuais respondem bem a públicos técnicos e regulatórios, mas traduzir essas explicações em termos compreensíveis para um cliente de varejo ou um conselho de administração continua sendo um desafio a ser superado.

33 Management Solutions (2023).
34 EBA (2021).
35 Management Solutions (3T23).

Vibe coding e engenharia de software aumentada

Da programação tradicional ao diálogo cognitivo com as máquinas

Historicamente, o desenvolvimento de software evoluiu por meio de saltos de abstração: da programação em assembly para linguagens de alto nível, do código imperativo para frameworks declarativos, do desenvolvimento manual para plataformas low-code. Cada transição eliminou a complexidade técnica desnecessária e aproximou a criação de software da intenção humana.

A IA generativa representa um salto qualitativo distinto: ela transforma a programação em uma conversa iterativa com sistemas cognitivos. O programador não escreve mais o código linha por linha: ele descreve o comportamento desejado em linguagem natural, e o sistema o gera, testa, corrige e documenta. Esse fenômeno, chamado de "*vibe coding*" por Andrej Karpathy[36], redefine o que significa programar: o desenvolvedor deixa de escrever sintaxe e passa a dirigir sistemas cognitivos que materializam a intenção em software funcional. Nas palavras de Karpathy, "o *vibe coding* transformará o software e alterará as descrições de cargos"[37].

O que é realmente o vibe coding?

O *Vibe coding* não é simplesmente o preenchimento automático de códigos sofisticados; é o desenvolvimento de software por meio da interação iterativa de linguagem natural com modelos de IA que mantêm a memória, o contexto e a compreensão de objetivos de alto nível.

Seus principais componentes incluem:

- Compreensão semântica do problema: o sistema interpreta os requisitos expressos em linguagem natural e os traduz em arquitetura técnica.
- Geração de código multiarquivo: o sistema produz não apenas fragmentos, mas aplicativos completos com estrutura modular coerente.
- Execução assistida, depuração e refatoração: o sistema executa o código, detecta erros, propõe correções e otimiza as implementações.
- Produção automática de testes e documentação: o sistema gera automaticamente baterias de testes e documentação técnica sincronizada com o código.

A diferença em relação aos assistentes de código tradicionais é fundamental: essas linhas ou funções são concluídas; os sistemas de vie coding entendem os objetivos, mantêm a coerência arquitetônica durante sessões prolongadas e operam como colaboradores cognitivos, não como ferramentas passivas.

Aceleração e democratização do desenvolvimento de software

O impacto na velocidade do desenvolvimento é quantificável e massivo. Um estudo de campo[38] com 4.867 desenvolvedores documentou um aumento de 26% na taxa de conclusão de tarefas. Um experimento[39] com 187.489 desenvolvedores mostrou um aumento de 12,4% no tempo gasto em atividades essenciais de programação, acompanhado por uma redução de 24,9% no tempo gasto em gestão de projetos e tarefas administrativas. Em outras palavras: projetos que costumavam levar meses agora são concluídos em semanas ou dias.

Outro efeito disruptivo é o equalizador: a IA reduz a diferença de produtividade entre desenvolvedores juniores e seniores. Enquanto os desenvolvedores juniores experimentam[40] ganhos de produtividade de 21% a 40%, os desenvolvedores sênior melhoram de 7% a 16%. Isso não significa que a experiência não seja mais importante, mas que o gargalo do desenvolvimento está mudando: não é mais essencial dominar a sintaxe, mas sim entender os problemas, projetar arquiteturas robustas e formular restrições com precisão.

Essa compreensão da lacuna de habilidades tem uma consequência organizacional direta: a profunda democratização da criação de software. Analistas de negócios, gerentes de produtos, consultores e cientistas estão gerando protótipos funcionais sem a intermediação de equipes de engenharia. O software não é mais exclusividade de perfis técnicos especializados. A barreira de entrada não é mais o conhecimento de linguagens de programação, mas a capacidade de articular problemas, objetivos e restrições com clareza.

O resultado líquido é uma queda estrutural no custo marginal da criação de software, levando a uma mudança de regime econômico na produção de tecnologia.

Transformação das equipes de tecnologia

Nas organizações de tecnologia, a distribuição de funções está mudando rapidamente. As equipes precisam de menos "programadores de baixo nível" escrevendo códigos de rotina e mais arquitetos de sistemas, designers de soluções, validadores de qualidade e auditores de riscos técnicos. A função dos desenvolvedores sênior está evoluindo: eles passam menos tempo na sintaxe e mais na governança da arquitetura, segurança, escalabilidade e gerenciamento de dívida técnica.

A pesquisa mostra[41] que as equipes assistidas por IA exigem, em média, 79,3% menos colaboradores por projeto, sem sacrificar a complexidade técnica. Pequenas equipes estão produzindo sistemas de escala e sofisticação antes reservados a grandes departamentos de engenharia. Além disso, a exploração de novas tecnologias aumentou 21,8%, o que sugere que os desenvolvedores estão liberando a capacidade cognitiva para aprender, experimentar e expandir suas capacidades técnicas.

36 Andrej Karpathy (n. 1986), ex-diretor de IA da Tesla e ex-pesquisador sênior da OpenAI, com contribuições importantes em deep learning, visão computacional e sistemas autônomos.
37 Business Insider (2025b).
38 Stanford (2025).
39 Ibid.
40 Ibid.
41 Ibid.

Qualidade, dívida técnica e risco

No entanto, a velocidade tem custos ocultos. A geração de código é instantânea; a manutenção, a depuração e o dimensionamento continuam difíceis. O *vibe coding* introduz novos vetores de risco:

- Fragilidade oculta: o código gerado pode funcionar superficialmente, mas contém ineficiências, vulnerabilidades ou dependências frágeis que só se manifestam na produção sob carga ou em casos extremos. A isso se soma um risco anterior na cadeia: as especificações ambíguas ou mal formuladas, que antes um desenvolvedor sênior teria questionado ou reinterpretado com bom senso, agora são executadas literalmente sem qualquer correção, propagando o erro silenciosamente desde o requisito até o produto.
- Dependência do modelo: se o sistema de IA que gerou o código desaparecer ou mudar significativamente, a capacidade de manter ou estender o software será prejudicada.
- Bugs sistêmicos replicados em escala: um bug no *prompt* ou na lógica do modelo pode se propagar instantaneamente para dezenas de projetos, multiplicando o impacto de bugs que antes seriam isolados.

A natureza do débito técnico também está mudando. Antes, a dívida técnica era principalmente dívida de código: implementações rápidas, refatoração adiada, duplicação de lógica. Agora, o débito técnico inclui dívidas de arquitetura (decisões de design implícitas nas interações com a IA), dívidas de *prompt*s (instruções mal formuladas que geram código abaixo do ideal, mas funcional) e dívidas de rastreabilidade (perda de compreensão do motivo pelo qual o código faz o que faz).

Governança de software na era do vibe coding

Se qualquer pessoa pode criar software por meio de conversas com IA, o risco organizacional se multiplica. O problema não é mais apenas saber qual código existe, mas qual sistema cognitivo o produziu, sob quais instruções e com qual grau de autonomia.

As principais estruturas de segurança e desenvolvimento já estão alertando sobre essa mudança. A OWASP identifica[42] novos riscos estruturais em aplicativos baseados em LLM, como *prompt* injection, insecure output handling e excessive agency: dar a um sistema de IA a capacidade de agir sem controles suficientes.

Ao mesmo tempo, o NIST insiste[43] que os princípios clássicos do desenvolvimento seguro (rastreabilidade, revisão, teste, controle de alterações, validação contínua) também devem ser aplicados ao conteúdo gerado pela IA e aos mecanismos que o transformam em alterações executáveis.

Como resultado, a governança de software não é mais apenas governança de código, mas governança de sistemas cognitivos, o que exige a introdução de novas camadas de controle:

- Repositórios de *prompt*s: instruções de catálogo, versão e auditoria que geram código crítico, pois o *prompt* se torna uma superfície de risco.
- Gestão de versões de intenção: registrar não apenas qual código foi produzido, mas também qual objetivo foi perseguido e quais restrições foram definidas.
- Controle de autonomia e permissões: limitar explicitamente as ações que a IA pode executar (modificação de repositório, implementações, comandos, acesso a dados).
- Rastreabilidade das decisões de projeto: documentar quais decisões foram tomadas por humanos e quais foram propostas ou executadas pela IA.
- Validação e revisão assistida: revisão sistemática de diffs, testes automatizados e auditorias de comportamento antes de aceitar alterações em produção.

As próprias ferramentas de desenvolvimento baseadas em agentes já recomendam explicitamente essas proteções (revisão de alterações, controle de permissões, cautela com a execução automática), refletindo que o risco não é teórico: a superfície de ataque e falha mudou do código isolado para o ciclo completo de intenção → geração → execução.

Nesse novo ambiente, as políticas de desenvolvimento não podem mais se limitar a padrões de codificação e processos de revisão. Elas devem incorporar regras de uso de IA, definição de limites de autonomia e protocolos de validação para resultados gerados automaticamente. A governança de software torna-se, pela primeira vez, governança de inteligência operacional.

Implicações estratégicas

O *vibe coding* não é apenas uma tendência tecnológica; é uma variável macroeconômica. As organizações que dominam esse recurso operam com vantagens competitivas estruturais: tempo de colocação no mercado extremamente reduzido, experimentação massiva de baixo custo e adaptabilidade organizacional acelerada.

Para as empresas tradicionais, a implicação é clara: elas estão competindo com organizações que iteram dez vezes mais rápido, com equipes dez vezes menores e com custos marginais de desenvolvimento que tendem a zero. A velocidade de criação de software deixa de ser uma métrica operacional interna e passa a ser um determinante da sobrevivência competitiva.

42 OWASP (2025).
43 NIST (2024a).

IA agêntica e sistemas autônomos

De assistentes de conversação a operadores autônomos

A IA generativa transformou o trabalho intelectual ao permitir que os profissionais gerassem conteúdo, analisassem informações e obtivessem respostas por meio de conversas naturais. No entanto, esses sistemas permanecem fundamentalmente reativos: eles respondem a solicitações, mas não agem de forma independente em sistemas reais. A IA agêntica representa um salto quântico: sistemas que planejam, executam tarefas complexas, tomam decisões dentro de limites definidos e operam em infraestruturas corporativas com rastreabilidade total.

Um sistema agêntico opera por meio de agentes autônomos, cada um com recursos específicos (raciocínio, memória, uso de ferramentas, planejamento), que colaboram para atingir um objetivo definido. Os recursos incrementais da IA agêntica em relação à IA generativa são estruturais[44]:

- **Status e memória**: mantém um contexto persistente entre as interações, não apenas em uma conversa isolada.
- **Planejamento dinâmico**: decompõe metas complexas em subtarefas, prioriza-as e replaneja de acordo com os resultados.
- **Execução em sistemas reais**: usa ferramentas e APIs para modificar dados, executar comandos e concluir transações.
- **Orquestração multiagente**: coordena agentes especializados sob um coordenador central que gerencia dependências e transferência de informações.
- **Rastreabilidade completa**: produz registros, evidências e justificativas de cada etapa executada, permitindo auditoria e supervisão.

IA agêntica em produção

A IA agêntica opera hoje em organizações globais que gerenciam milhões de transações diariamente. Para citar alguns exemplos:

- **Deutsche Bank**: está implantando um agente habilitado para voz de IA ("AI banking butler") que atua como um agente de conversação proativo, desde o atendimento e suporte até a execução de transações e serviços de consultoria[45]. O programa faz parte de um investimento em tecnologia de aproximadamente 600 milhões de euros, com uma meta de economia recorrente de cerca de 300 milhões de euros por ano até 2028, e é acompanhado por uma redução de cerca de 10% na força de trabalho[46].
- **Ryt Bank**: banco malaio que se anuncia como "AI-first", opera em uma arquitetura de agentes especializados em que o cliente interage em linguagem natural e os agentes interpretam a intenção, orquestram processos e executam transações reais no core banking. O sistema lida com cerca de 80.000 transações por mês, reduziu processos que exigiam de 5 a 8 telas para uma única interação de conversação e demonstrou uma melhoria significativa na retenção de clientes[47].

44 iDanae (2T25).
45 Deutsche Bank (2025).
46 Financial News London (2025).
47 Ryt Bank (2025).

- **Walmart**: em seus centros de distribuição automatizados, os sistemas de decisão autônomos coordenam o armazenamento, o reabastecimento e a coleta de pedidos em tempo real. O Walmart informa que esses centros dobram a capacidade de processamento com aproximadamente metade da equipe em comparação com os centros tradicionais, demonstrando uma mudança estrutural na produtividade logística[48].
- **Amazon**: seu novo sistema de gerenciamento de logística, o Sequoia, orquestra robôs, estoque e fluxos de pedidos por meio de um software autônomo que decide onde armazenar, quando mover o estoque e como estocar as estações de coleta. A Amazon relata reduções de até 75% no tempo de manuseio do estoque e de até 25% no tempo de processamento de pedidos nos centros onde está implantado[49].
- **DHL**: Em operações de separação com robôs autônomos integrados ao WMS, os sistemas atribuem trabalho automaticamente, otimizam rotas e encerram tarefas. Em implantações produtivas, a DHL relatou aumentos de produtividade de até +180% em unidades por hora, juntamente com melhorias significativas na qualidade e na precisão[50].

48 Business Insider (2025a).
49 Amazon (2023).
50 DHL (2024).

Mas o que é um sistema agêntico? Cinco camadas modulares

Estas capacidades funcionais se concretizam em uma arquitetura modular específica (não simplesmente um modelo de linguagem com acesso a ferramentas), composta de cinco camadas interdependentes:

1. **Interface e percepção:** recebe metas do usuário, fornece resultados finais e percebe eventos do ambiente por meio de gateways de API, pontos de extremidade e conectores de entrada.
2. **Orquestração e agendamento**: decompõe metas complexas em tarefas gerenciáveis, decide qual agente executa cada tarefa e em que ordem, e gerencia o enfileiramento e o roteamento de prioridades.
3. **Núcleo do agente**: trabalhadores autônomos, cada um com uma função específica, núcleo cognitivo (LLM) e loop de controle ReAct (Reason + Act) que combina raciocínio e ação iterativa.
4. **Ferramentas e serviços**: biblioteca de recursos externos (pesquisa, geração de código, APIs corporativas) com conectividade padronizada por meio de protocolos como MCP e gerenciamento de *prompt*s.
5. **Memória e conhecimento**: armazena informações de curto prazo (histórico de conversas), informações de longo prazo (banco de dados de vetores com experiências passadas) e base de conhecimento corporativo.

Essa arquitetura torna a autonomia rastreável, auditável e governável. Cada decisão, cada ação, cada invocação de ferramentas é registrada, permitindo uma supervisão humana eficaz e o compliance regulatório.

MCP: o elo perdido para a escalabilidade

A arquitetura agêntica enfrenta um problema técnico fundamental: a complexidade exponencial das integrações. Tradicionalmente, cada modelo de linguagem requer uma integração proprietária com cada ferramenta. A alteração de modelos requer a reescrita de todas as integrações; a adição de uma nova ferramenta requer a integração com todos os modelos existentes. O resultado é uma dívida técnica exponencial.

O Model Context Protocol (MCP) resolve esse problema por meio de uma camada de abstração universal. O MCP é um padrão aberto que padroniza a forma como os modelos de IA interagem com aplicativos externos, fontes de dados e ferramentas. Os "servidores MCP" expõem recursos (recursos, *prompt*s, ferramentas) que os "clientes MCP" (agentes ou modelos) consomem sob demanda. Uma ferramenta conectada à MCP é imediatamente acessível a qualquer agente atual ou futuro, sem a necessidade de redesenvolver integrações.

O impacto do MCP é transformador: ele passa de integrações feitas à mão para ativos universalmente reutilizáveis, facilita a transição de protótipos isolados para ecossistemas dimensionáveis, permite que os agentes adquiram novos recursos sem reimplantação e reduz drasticamente o custo de manutenção e evolução. Sem o MCP ou padrões equivalentes, a IA agêntica em escala empresarial não é tecnicamente sustentável.

Governança e controle: o verdadeiro desafio

Criar agentes é relativamente simples com as estruturas atuais; governá-los em escala empresarial é o verdadeiro desafio. A adoção sustentável exige o equilíbrio de quatro recursos:

- **Tecnologia e desenvolvimento**: orquestração de vários agentes com coordenação eficaz, gerenciamento de memória operacional e de longo prazo, arquiteturas modulares que permitem manutenção e evolução e integração segura com sistemas corporativos por meio de padrões como o MCP.
- **Avaliação contínua**: métricas de desempenho, qualidade e eficiência, monitoramento de desvios e anomalias, controle de custos (as chamadas para modelos se multiplicam exponencialmente) e rastreabilidade total de ações e decisões.
- **Governança e compliance**: supervisão humana por meio de mecanismos de aprovação e escalonamento, (levando em conta que a capacidade humana de supervisão tem um limite; uma vez ultrapassado esse limite, a supervisão torna-se meramente formal e gera uma falsa sensação de controle) controles explícitos e limites sobre as ações que cada agente pode executar, transparência e explicabilidade funcional das decisões e compliance com o *AI Act* e as políticas internas de risco.
- **Industrialização e modelo operacional**: operação 24 horas por dia, 7 dias por semana, com manutenção contínua, pipelines de CI/CD para implantar e atualizar agentes com segurança, gerenciamento ativo de custos na produção e resiliência a falhas ou comportamentos inesperados.

Implicações estratégicas

A adoção da IA agêntica tem três implicações estratégicas fundamentais:

- **Transformação do modelo operacional**: a IA agêntica introduz uma mudança estrutural em que as equipes humanas colaboram com agentes autônomos que operam 24 horas por dia, 7 dias por semana, sem supervisão contínua. Ela reduz radicalmente o tempo de colocação no mercado: tarefas que costumavam levar semanas agora são resolvidas em dias por meio da delegação a agentes especializados.
- **Risco de aumento de custos**: ao contrário da IA generativa tradicional, os agentes multiplicam as chamadas para modelos e ferramentas. Um protótipo viável pode se tornar um sistema financeiramente insustentável se não for projetado com controles de custo desde o início.
- **Industrialização como barreira de entrada**: a criação de protótipos agênticos é possível com as estruturas atuais, mas operá-los, dimensioná-los, mantê-los, protegê-los e auditá-los na produção exige recursos organizacionais que a maioria das empresas ainda não desenvolveu. A vantagem competitiva não está na tecnologia, mas na capacidade de industrializá-la com uma governança eficaz.

IA em robótica e sistemas físicos

Do digital à ação no mundo real

Durante anos, a robótica industrial operou por meio de sistemas programados para tarefas repetitivas em ambientes altamente controlados: braços robóticos que montam componentes seguindo sequências fixas, veículos guiados automaticamente (AGVs) que seguem rotas predefinidas, sistemas de "pick and place" que reconhecem objetos em posições exatas. Esses robôs executam movimentos precisos, mas não têm adaptabilidade: qualquer mudança no ambiente (um objeto mal posicionado, uma variação na textura, um obstáculo inesperado) exige reprogramação ou intervenção humana.

A integração de IA generativa, modelos avançados de visão e aprendizagem por reforço está transformando essa realidade no setor: somente em 2023, mais de 276.000 robôs industriais foram instalados na China (Fig. 2), o que já representa mais da metade das instalações mundiais, e a participação de robôs colaborativos quadruplicou em seis anos[51]. Os robôs atuais percebem seu ambiente por meio de visão computacional em tempo real, interpretam instruções em linguagem natural, planejam sequências complexas de ações, adaptam-se a mudanças imprevistas sem reprogramação e aprendem com cada interação para melhorar continuamente. A IA transforma máquinas industriais rígidas em sistemas autônomos capazes de operar em ambientes não estruturados e executar tarefas que antes exigiam inteligência humana.

51 Stanford (2025).

Robôs humanoides: do laboratório para o chão de fábrica

A robótica humanoide deu um salto quântico nos últimos dois anos, passando de demonstrações espetaculares em laboratório para implementações industriais reais.

A Figure AI, uma startup avaliada em aproximadamente US$ 39 bilhões, concluiu uma implantação de 11 meses de seus robôs Figure 02 na fábrica da BMW em Spartanburg (Carolina do Sul) em 2025[52]. Os dois robôs humanoides trabalharam em turnos de 10 horas de segunda a sexta-feira, acumulando 1.250 horas de operação, carregando mais de 90.000 peças de chapa metálica com tolerâncias de 5 milímetros em 2 segundos por peça e contribuindo para a produção de mais de 30.000 veículos BMW X3.

A Figure lançou sua terceira geração, a Figure 03, projetada especificamente para produção em volume. A empresa construiu a BotQ, uma instalação dedicada à fabricação de robôs humanoides com capacidade inicial de 12.000 unidades por ano e uma meta de produzir 100.000 robôs em quatro anos. A Figure 03 incorpora carregamento indutivo sem fio (2 kW por meio de bobinas de pé que permitem que o robô simplesmente pise em uma base para recarregar), um sistema de visão reprojetado com o dobro da taxa de atualização e um quarto do tempo de latência, além de câmeras de palma integradas para feedback visual redundante durante a manipulação fina. A empresa projeta que esses robôs operarão usando sua IA proprietária Helix, treinada em massa com dados de teleoperação e demonstrações humanas.

Por sua vez, a Tesla está aumentando agressivamente a produção de seu robô humanoide Optimus. A empresa anunciou planos para construir uma linha de produção capaz de fabricar um milhão de unidades por ano, com início previsto para o final de 2026[53]. Elon Musk declarou em outubro de 2025 que a versão 3 do Optimus terá "mãos que são uma incrível obra de engenharia" com amplitude total de movimento humano (22 graus de liberdade) e que o robô será tão realista que "você precisará tocá-lo para acreditar que é realmente um robô"[54]. A Tesla produziu vários milhares de unidades até 2025 para uso interno em suas fábricas (principalmente para manuseio de baterias e componentes) e planeja aumentar para 50.000-100.000 unidades até 2026[55]. Musk estima[56] que, em volumes acima de 1 milhão de unidades por ano, o custo de produção do Optimus cairá abaixo de US$ 20.000, aproximadamente metade do custo de um Modelo Y em escala equivalente. O preço de varejo, no entanto, será significativamente mais alto e determinado pela demanda do mercado.

A Boston Dynamics, referência histórica em robótica dinâmica, aposentou em abril de 2024 seu Atlas hidráulico (famoso por backflips e parkour) e lançou um Atlas totalmente elétrico projetado para aplicações industriais reais[57]. O novo Atlas integra atuadores personalizados de alto desempenho com amplitude de movimento que excede a capacidade humana: sua cabeça e seu tronco podem girar 180 graus independentemente, suas articulações têm extrema flexibilidade e ele foi projetado para explorar sua própria anatomia mecânica, não se limitando a posturas humanas, embora

52 Figure (2025).
53 Teslarati (2025).
54 Tesla Car World (2025).
55 Fortune (2025b).
56 Notateslaapp (2025).
57 Boston Dynamics (2025a).

muitos de seus recursos de controle sejam treinados a partir do movimento humano. A Boston Dynamics enfatiza que o Atlas priorizará a velocidade e a eficiência em detrimento da aparência antropomórfica.

Em agosto de 2025, a Boston Dynamics e o Toyota Research Institute demonstraram o[58] Atlas operando por meio de Large Behavior Models (LBMs): políticas completas treinadas com demonstrações extensas e anotações de linguagem que coordenam a locomoção e a manipulação simultaneamente. Um único modelo de comportamento controla diretamente todo o robô, tratando mãos e pés de forma quase idêntica, sem separar o controle de locomoção de baixo nível do controle de manipulação. Em vídeos públicos[59], o Atlas executa sequências contínuas de tarefas de classificação e embalagem em ambientes simulados de fábrica, reagindo de forma autônoma a distúrbios físicos inesperados (como pesquisadores fechando caixas ou empurrando objetos) sem interromper a tarefa. A Boston Dynamics planeja pilotos[60] com a Hyundai em 2026 e produção comercial limitada a partir de 2027.

Além da manufatura e da logística, a robótica humanóide aponta para uma segunda área de impacto: o cuidado de pessoas idosas, dependentes ou com deficiência. Em economias com envelhecimento estrutural acelerado, esta aplicação tem relevância estratégica comparável à industrial, com implicações éticas e regulatórias próprias que os marcos atuais mal começaram a abordar.

Implicações estratégicas e desafios operacionais

A integração da IA à robótica humanoide introduz mudanças estruturais na fabricação e na logística. A produtividade aumenta drasticamente: os robôs operam 24 horas por dia, 7 dias por semana, sem fadiga, pausas ou variação de desempenho. Os custos operacionais recorrentes (energia, manutenção preventiva) são previsíveis e estão diminuindo com as economias de escala, embora o investimento inicial continue sendo significativo.

No entanto, surgem desafios críticos. O impacto sobre o emprego é real e concentrado: tarefas manuais repetitivas na fabricação, montagem e manuseio de materiais enfrentam uma automação acelerada. As organizações que adotam a robótica humanoide precisam gerenciar as transições de mão de obra, os programas de requalificação e as crescentes expectativas regulatórias e sociais sobre a responsabilidade corporativa, desafios que são comuns à adoção da IA em geral, mas que, neste contexto, assumem maior concentração setorial e visibilidade social

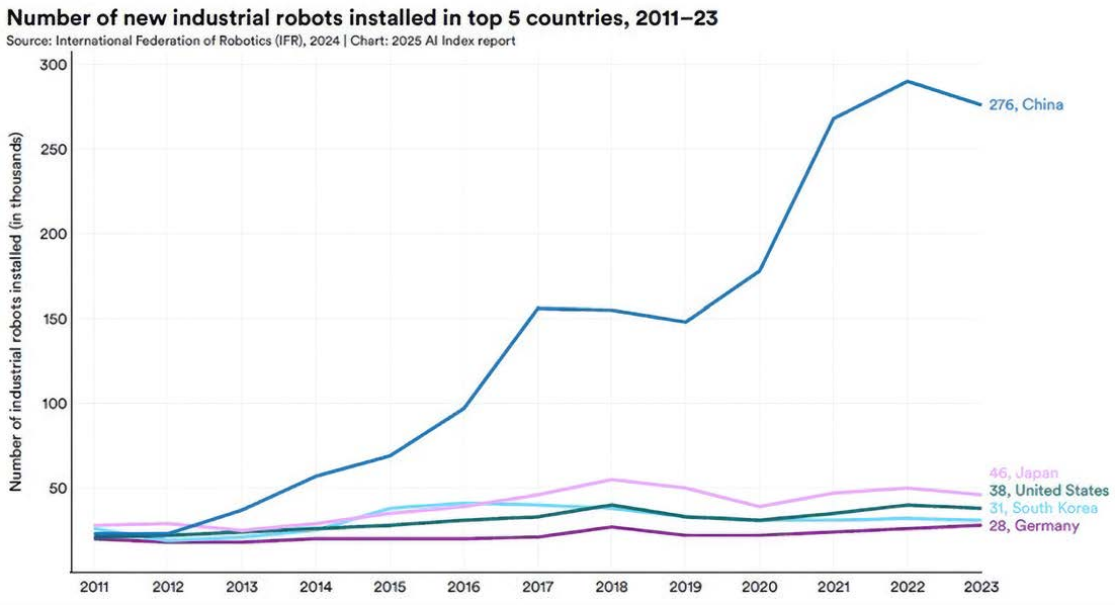
A dependência de fornecedores se intensifica: as empresas que adotam robôs estão vinculadas a seus ecossistemas proprietários de hardware, software, atualizações de IA e suporte técnico. A obsolescência tecnológica é rápida: um robô comprado hoje pode ser superado em termos de recursos pela próxima geração em dois anos, o que levanta questões sobre ciclos de investimento e estratégias de atualização.

E surgem novos riscos operacionais: sistemas autônomos que operam em ambientes físicos compartilhados com humanos podem causar ferimentos, danos à propriedade ou interrupções operacionais críticas se falharem. A robótica com IA exige estruturas de segurança robustas (sensores redundantes, sistemas de desligamento de emergência, zonas de exclusão dinâmicas), protocolos certificados à prova de falhas e supervisão humana eficaz, mesmo em operações nominalmente autônomas.

A IA na robótica não é uma tendência futura, é uma realidade operacional em uma fase de industrialização acelerada. As organizações que avaliarem estrategicamente quando e onde adotar a robótica humanoide (nem todas as tarefas justificam o investimento) obterão vantagens competitivas sustentáveis.

58 Boston Dynamics (2025b).
59 Toyota Research Institute (2025).
60 Hyundai (2025).

Fig. 2. Robôs industriais recém-instalados. Fonte: Stanford (2025).

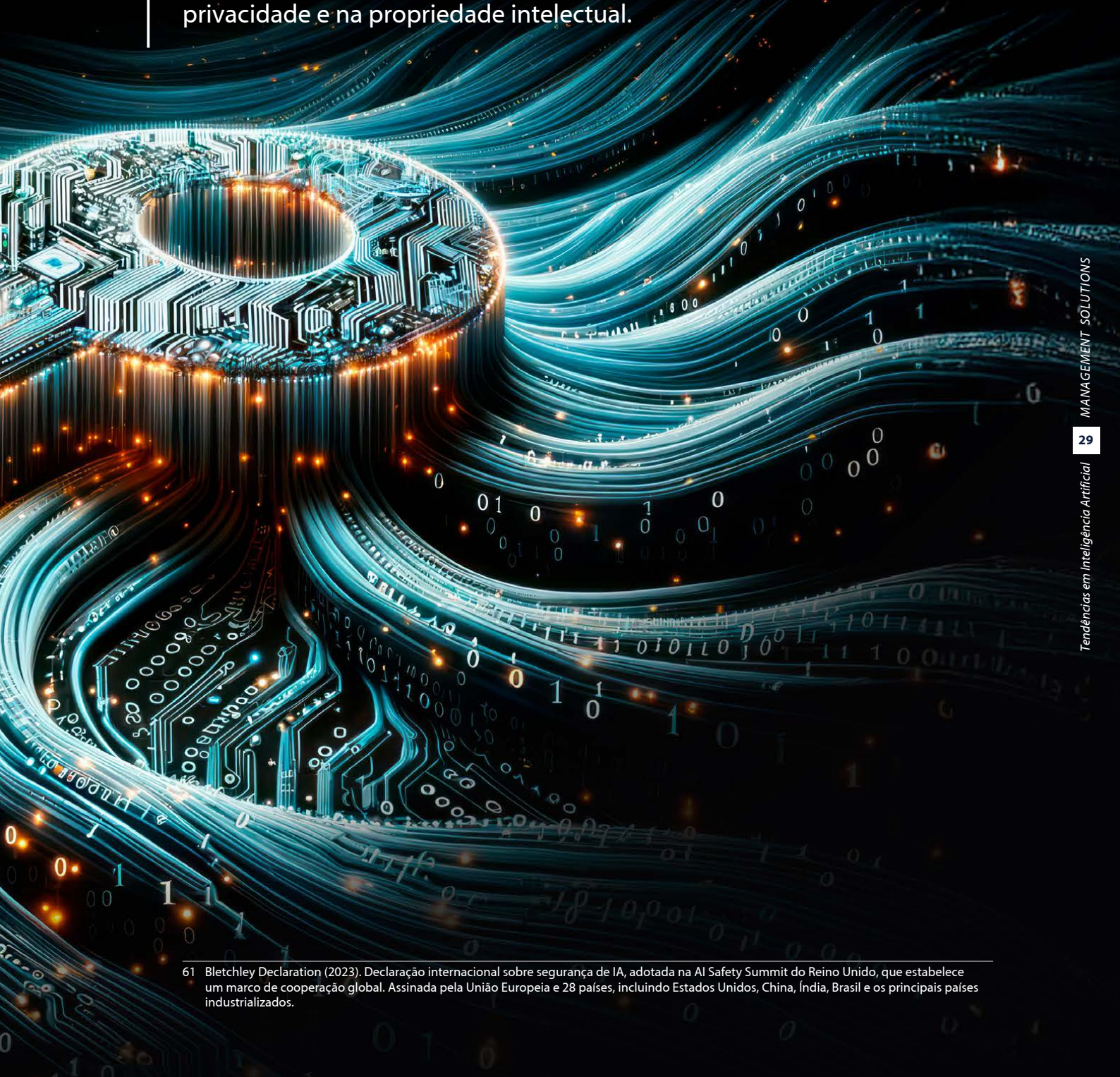


04 | Riscos, regulação e segurança da IA

«Há potencial para danos graves, até mesmo catastróficos, deliberados ou não, decorrentes das capacidades mais significativas desses modelos de IA».

Bletchley Declaration[57]

A IA generativa e os sistemas agênticos oferecem recursos transformadores, mas introduzem riscos estruturais que não são mais hipotéticos: eles estão se materializando na produção. A Declaração de Bletchley advertiu explicitamente sobre os impactos potencialmente "catastróficos" da IA, e a experiência acumulada confirma que a lista de riscos é longa e diversificada. Esta seção analisa os riscos da IA, as estruturas regulatórias e a governança da IA, a batalha entre a IA defensiva e adversária na segurança cibernética e as tensões abertas na privacidade e na propriedade intelectual.



61 Bletchley Declaration (2023). Declaração internacional sobre segurança de IA, adotada na AI Safety Summit do Reino Unido, que estabelece um marco de cooperação global. Assinada pela União Europeia e 28 países, incluindo Estados Unidos, China, Índia, Brasil e os principais países industrializados.

Riscos da IA

A IA não cria riscos: ela os amplifica

A adoção da IA não introduz, com algumas exceções, riscos substancialmente novos, mas amplia drasticamente os riscos existentes: operacionais, de modelo, tecnológicos, de fornecedores, jurídicos, de reputação, de compliance, estratégicos, sociais... A diferença não está na natureza do risco, mas em sua velocidade de propagação, escala de impacto e dificuldade de contenção.

Com exceção de algumas categorias emergentes (como a geração de conteúdo tóxico, certas formas de manipulação cognitiva por meio de *deepfakes* ou ataques de *prompt* injection), a maioria dos riscos associados à IA são versões aceleradas, automatizadas e massivas de problemas conhecidos. Uma tendência algorítmica é, em essência, uma tendência humana sistematizada e replicada milhões de vezes. Um vazamento de informações por meio do uso indevido de um chatbot é, no fim das contas, um vazamento de informações.

Quatro dimensões interconectadas

Na prática, esses riscos se materializam em quatro grandes dimensões (Fig. 3):

1. Segurança e compliance.

A IA introduz novas superfícies de ataque e complica a compliance regulatório. A privacidade e a segurança das informações enfrentam ameaças específicas: vazamentos não intencionais de dados confidenciais, vulnerabilidades técnicas emergentes, como *prompt* injection (instruções maliciosas incorporadas em entradas que "enganam" o modelo para que ele ignore as restrições) ou jailbreaks (técnicas para contornar os controles de segurança) e exposição acidental de informações confidenciais quando os profissionais usam ferramentas não seguras.

A propriedade intelectual torna-se ambígua: a quem pertence o código gerado pela IA, o que acontece quando um modelo reproduz fragmentos protegidos por direitos autorais? Como exemplo, o New York Times está em um litígio aberto com a OpenAI/Microsoft por usar conteúdo protegido por direitos autorais sem licença, entre muitos outros processos em andamento[62].

A rastreabilidade e a reprodutibilidade são prejudicadas quando decisões críticas dependem de modelos que evoluem continuamente por meio de retreinamento automático, dificultando a reconstrução exata de qual versão do sistema produziu qual resultado e em qual momento.

E o não compliance agora tem consequências financeiras diretas e visíveis: a Lei Europeia de IA[63] prevê multas de até 35 milhões de euros ou 7% do faturamento global anual, tornando o compliance com a IA um risco material da mais alta ordem, maior até do que a proteção de dados[64].

2. Qualidade, confiabilidade, *lock-in* e custos

Conforme mencionado acima, a IA generativa produz resultados plausíveis, não necessariamente corretos. As alucinações (geração de informações falsas apresentadas com confiança) não são erros ocasionais, mas um comportamento intrínseco ao projeto real dos modelos. O comportamento não determinístico significa que a mesma entrada pode produzir saídas diferentes no mesmo modelo, o que complica a validação, a auditoria e a certificação de processos críticos.

O desvio silencioso do modelo representa um dos riscos mais insidiosos: um modelo que funciona bem pode se degradar em um curto espaço de tempo porque a distribuição dos dados mudou, sem que existam mecanismos de monitoramento que gerem alertas antecipados Um modelo pode continuar a operar com uma aparência de normalidade até que alguém detecte manualmente anomalias nos resultados.

A dependência excessiva da IA em tarefas críticas cria fragilidade operacional: se o sistema de IA falhar, travar ou se tornar proibitivamente caro, a organização poderá continuar operando? Existem procedimentos humanos de contingência? E a perda do controle humano efetivo ocorre quando as decisões são delegadas a sistemas cujo raciocínio interno é opaco até mesmo para aqueles que os desenvolvem e operam.

Além disso, as organizações criam dependências estruturais em alguns poucos provedores de modelos (OpenAI, Anthropic, Google) e infraestrutura (AWS, Azure, GCP). Mudanças nos preços, nos termos de serviço ou interrupções operacionais podem prejudicar processos essenciais simultaneamente. A migração entre ecossistemas envolve reescrever integrações, retreinar workflows, recertificar o compliance e assumir custos proibitivos, o que nem sempre é viável, a diversificação de fornecedores, embora dispendiosa, é a resposta estrutural a este risco

Além dos riscos acima, há uma dimensão econômica que os frameworks de gestão tradicionais subestimam. Os custos unitários de inferência vêm caindo estruturalmente (aproximadamente 10 vezes a cada 12 meses), mas essa queda não protege contra o crescimento exponencial do volume: os sistemas agênticos têm estruturas de custo com escalabilidade não linear: cada agente multiplica as chamadas para modelos e ferramentas e, sem monitoramento de tokens em tempo real e limites de gastos explícitos desde o desenho, um protótipo viável pode se tornar um sistema insustentável antes que alguém perceba[65]. Soma-se a isso a incerteza quanto ao retorno sobre o investimento: a questão de saber se os gastos massivos com IA gerarão o valor esperado ainda não está resolvida, e muitas organizações avançam impulsionadas pela pressão da concorrência em vez de um caso comercial sólido. O Gartner sintetiza os dois riscos em uma previsão: mais de 40% dos projetos de agentic serão cancelados até 2027 devido ao aumento dos custos, ao valor comercial pouco claro e aos controles de risco inadequados[66]. O lock-in agrava o problema estruturalmente: a dependência de um único fornecedor elimina a capacidade de migrar para arquiteturas mais eficientes quando elas estiverem disponíveis.

3. Ética e tomada de decisão automatizada

Os vieses algorítmicos amplificam os vieses presentes nos dados históricos: se um modelo de recrutamento for treinado com dados

62 Brown (2025).
63 Lei de IA (2024).
64 Management Solutions (4T24).

65 A estimativa de custos antes do desenvolvimento e o monitoramento contínuo em produção são duas capacidades que as organizações estão começando a formalizar em suas estruturas de governança de IA. O GenMS™ Tracker, solução proprietária da Management Solutions, aborda ambas as dimensões: permite estimar o custo de desenvolvimento e operação de um sistema de IA a partir de uma descrição em linguagem natural e monitora em tempo real o consumo em produção com alertas em caso de desvios em relação ao orçamento definido..
66 Gartner (2025b).

Fig. 3. Alguns riscos da IA generativa.

de uma organização historicamente tendenciosa em relação a determinados perfis, o modelo sistematizará e aumentará esse viés[67]. A falta de transparência e as dificuldades de explicação complicam o compliance com os requisitos regulatórios de que as decisões automatizadas devem ser compreensíveis e justificáveis, especialmente quando afetam os direitos fundamentais.

As decisões automatizadas com impacto humano direto (aprovação de crédito, diagnósticos médicos, avaliação de risco criminal, seleção de candidatos) levantam questões de responsabilidade: quem é responsável quando o modelo comete um erro com consequências graves: o fornecedor do modelo básico, a equipe que o personalizou, o usuário que escreveu o *prompt*, o comitê que aprovou sua implantação? A cadeia de responsabilidade torna-se complexa, criando lacunas de responsabilidade que nem a regulação nem a prática resolveram completamente.

4. Impacto social e organizacional

A transformação do trabalho não é mais especulativa: as tarefas manuais e cognitivas de rotina enfrentam uma automação acelerada, e as organizações precisam gerenciar as transições de trabalho, os programas de requalificação e as crescentes expectativas regulatórias e sociais sobre a responsabilidade corporativa.

A dependência cognitiva e a erosão de recursos essenciais representam um grande risco: se toda uma geração de profissionais aprender a trabalhar exclusivamente com IA como intermediária, eles manterão a intuição e o conhecimento suficientes para avaliar criticamente os resultados? O que acontecerá quando a IA não estiver disponível? E a pressão sobre os recursos e a pegada ambiental não é trivial: o treinamento de modelos avançados consome energia equivalente a milhares de residências durante meses, e a inferência em escala massiva levanta questões sobre a sustentabilidade de longo prazo[68].

Amplificação não linear

A IA introduz um fenômeno fundamental na gestão de riscos: a amplificação não linear. Uma pequena falha (um viés de dados, um *prompt* mal projetado, uma configuração incorreta de permissão) pode aumentar em minutos e afetar simultaneamente os processos, os clientes, os órgãos reguladores e a reputação. Quanto maior for a autonomia e a integração da IA nos processos essenciais, maior será o raio de impacto da falha, o que exige a concepção, desde o início, de mecanismos de contenção proporcionais a essa escala.

Um exemplo concreto: um modelo de atendimento ao cliente que, após uma pequena alteração no *prompt* do sistema, começa a revelar informações confidenciais de outros clientes em "apenas" 0,01% das conversas, o que é quase indetectável nos testes. Em um sistema que lida com 100.000 interações por dia, isso significa 10 vazamentos por dia antes que o padrão seja detectado. Quando o problema é identificado, centenas de incidentes já ocorreram, cada um com implicações regulatórias, contratuais e de reputação.

IA na taxonomia de riscos corporativos

As organizações estão adotando uma das três abordagens para colocar o risco de IA em sua estrutura de risco corporativo:

- Tratá-lo como um risco de nível superior, criando uma nova categoria na taxonomia (uma abordagem muito rara, mas útil para obter visibilidade executiva nos estágios iniciais da adoção em massa).
- Tratá-lo como um risco de segundo nível, vinculado ao risco tecnológico, de modelo ou operacional, dentro da taxonomia existente.
- Tratá-lo não como uma categoria independente, mas como um fator transversal que amplia os riscos existentes (risco de modelo, risco de fornecedor, risco de reputação, etc., amplificado pela AI).

Na prática, o rótulo é menos importante do que a capacidade da organização de identificar, prevenir, controlar e mitigar esses riscos de forma sistemática e contínua, com métricas claras, responsabilidades atribuídas e mecanismos de escalonamento definidos.

Implicações estratégicas

O gerenciamento de riscos de IA tornou-se uma decisão estratégica que condiciona a velocidade de adoção, a viabilidade operacional e a credibilidade institucional da organização.

As organizações que abordam a IA principalmente como um projeto de inovação tecnológica enfrentam atritos com a estrutura regulatória, incidentes operacionais e tensões de reputação. A integração da IA desde a sua concepção na arquitetura de governança, controle interno e gestão de riscos permite o dimensionamento com maior estabilidade, menos atrito e maior confiança do órgão regulador, do mercado e dos próprios profissionais.

A vantagem competitiva sustentável decorre de uma maior capacidade de governar a IA: estruturas de controle bem definidas, responsabilidades explícitas, monitoramento contínuo e cultura organizacional que preserva o julgamento humano no centro das decisões críticas.

67 Management Solutions (2023).
68 iDanae (1T24).

Regulação, supervisão e padrões de IA

IA, uma atividade regulada

A adoção acelerada da IA e a percepção de seus riscos desencadearam uma resposta regulatória global sem precedentes. Diferentemente dos ciclos tecnológicos anteriores, em que a regulação reagiu com anos de atraso, a IA está sendo regulada paralelamente à sua implantação em massa, justamente porque seus riscos sistêmicos já estão se materializando.

A Europa assumiu a liderança regulatória com o *AI Act*[69], a primeira estrutura legal abrangente sobre IA. Iniciativas relevantes nos Estados Unidos, na China, no Reino Unido, no Canadá e em outros países se seguiram, juntamente com um ecossistema crescente de padrões técnicos e estruturas voluntárias que buscam operacionalizar princípios de governança, segurança e ética[70].

A consequência para as organizações é clara: o gerenciamento da IA não é mais uma questão tecnológica, mas um domínio regulatório estrutural, comparável em impacto à proteção de dados, aos mercados financeiros ou à segurança operacional.

O AI Act: a regulação mais prescritiva

O Regulamento Europeu de IA (Regulamento (UE) 2024/1689) estabelece um modelo regulatório baseado em risco, estruturando obrigações de acordo com o impacto potencial dos sistemas sobre os direitos fundamentais, a segurança e a ordem pública.

A estrutura classifica os sistemas de IA em quatro categorias principais:

1. Risco inaceitável: sistemas proibidos (manipulação cognitiva, pontuação social, certas formas de vigilância biométrica).
2. Alto risco: sistemas usados em áreas críticas, como crédito, emprego, educação, infraestrutura, justiça, saúde, biometria, entre outros. Esse é o núcleo do *AI Act*.
3. Risco limitado: sistemas que exigem obrigações de transparência.
4. Risco mínimo: sistemas que são livres para uso sob alguns princípios gerais.

Para sistemas de alto risco, o *AI Act* introduz um conjunto de obrigações estruturais:

- Sistema de gestão de riscos documentado.
- Governança de dados e controle de qualidade de dados.
- Documentação técnica abrangente.
- Registro e rastreabilidade.
- Supervisão humana efetiva.
- Requisitos de precisão, robustez e segurança cibernética.
- Avaliação de compliance prévia à implantação e vigilância contínua pós-comercialização.

69 *AI Act* (2024).
70 Management Solutions (4T24a).

As sanções por descumprimento grave do *AI Act* chega a 35 milhões de euros ou 7% do faturamento global anual, excedendo até mesmo o regime do GDPR (que é de 4%).

Supervisão na Europa: da inovação ao controle permanente

O *AI Act* cria uma nova arquitetura institucional:

- AI Office: o órgão técnico central da Comissão para IA, especialmente o GPAI.
- Autoridades supervisoras nacionais: supervisionam e sancionam o compliance com o *AI Act* em cada país.
- AI Board: fórum para coordenação e interpretação comum entre a Comissão e as autoridades nacionais.
- Mecanismos de cooperação transfronteiriça: regras para a coordenação de casos e investigações envolvendo vários Estados-Membros.

A supervisão não se limitará a auditorias pontuais: é estabelecido um modelo de supervisão contínua, com obrigações de relatório, gerenciamento de incidentes graves, retirada de sistemas inseguros e poderes de inspeção comparáveis aos dos reguladores financeiros.

Na prática, a arquitetura de supervisão do *AI Act* ainda está em construção. O AI Board, órgão central de coordenação entre os Estados-Membros e a Comissão, realizou sua primeira reunião formal[71] em setembro de 2024, e se concentrou em questões organizacionais, códigos de prática para modelos de AIFM e coordenação de autoridades nacionais. As atas revelam um processo ainda em fase de organização: seleção de um presidente, criação de subgrupos e discussão sobre resultados prioritários.

Por sua vez, as autoridades nacionais de supervisão quase não foram designadas. A Espanha criou a AESIA (Agencia Española de Supervisión de la Inteligencia Artificial)[72] em agosto de 2023, tornando-se a primeira autoridade de IA da Europa, e iniciou suas operações efetivas em fevereiro de 2025 com funções de supervisão de sistemas proibidos. Outros Estados-Membros estão em estágios anteriores de designação de autoridades.

71 AI Board (2026).
72 AESIA (2026).

O ecossistema de normas: dos princípios à operação

Paralelamente à regulação, uma rede de padrões técnicos e de gerenciamento está se consolidando para estabelecer normas para a prática operacional de IA:

- ISO/IEC 42001: estabelece requisitos para um sistema de gerenciamento de IA (AIMS), no estilo da ISO 27001/9001, para governança e uso responsável da IA.
- ISO/IEC 23894: proporciona um framework detalhado de gestão de riscos de IA ao longo do ciclo de vida, destinado a ser integrado com frameworks de gestão de riscos corporativos.
- ISO/IEC 5259: estrutura e métricas de qualidade de dados de IA.
- NIST AI Risk Management Framework: estrutura focada no gerenciamento de riscos de IA, organizada nas funções Govern-Map-Measure-Manage e orientada para recursos de "trustworthy AI ".
- OECD AI Principles: princípios de alto nível (valores humanos, transparência, robustez, responsabilidade, inclusão) que influenciaram diretamente o AI Act e outros marcos regulatórios.

Essas normas têm uma função essencial: estabelecem controles concretos, processos auditáveis e métricas verificáveis. Para muitas organizações, elas estão formando a base técnica de seus programas de compliance.

Fragmentação geopolítica e complexidade operacional

A abordagem europeia baseada em riscos e vinculante não está sendo replicada globalmente:

- **Os EUA** não têm uma estrutura federal; têm uma abordagem setorial fragmentada por agências (FTC, FDA, EEOC), sem equivalente ao *AI Act*, complementada por ordens executivas e orientações, além de regulação em vários estados (por exemplo, Califórnia, Colorado, Connecticut, Utah)[73].
- **A China** articula a IA dentro de uma estratégia estatal de "soberania digital", com regras específicas sobre algoritmos, deepfakes e modelos generativos, forte controle de dados e licenciamento compulsório para determinados sistemas de alto impacto[74].
- **O Reino Unido** mantém uma abordagem pró-inovação sem uma lei horizontal de IA, contando com autoridades setoriais e princípios comuns de regulação de IA, com diretrizes e sandboxes regulatórias[75].
- **O Brasil** aprovou no Senado (dezembro de 2024) um projeto de lei estruturalmente semelhante ao *AI Act*: abordagem baseada em risco (proibido/alto/limitado/mínimo), obrigações específicas para sistemas de alto risco, multas de até R$ 50 milhões ou 2% do faturamento. O projeto de lei está pendente de aprovação na Câmara dos Deputados, com entrada em vigor prevista para um ano após a promulgação[76].
- **O México** carece de regulação: mais de 60 projetos de lei foram apresentados desde 2020 sem aprovação. Em fevereiro de 2025, uma reforma constitucional foi introduzida para conceder competência federal em IA, que deve ser traduzida em uma Lei Geral. Até lá, não há uma estrutura legal específica, apenas diretrizes voluntárias alinhadas com a UNESCO e a OCDE[77].
- **A Austrália** mantém uma abordagem voluntária e baseada em princípios: AI Ethics Principles (2019, 8 princípios voluntários) e Guidance for AI Adoption (outubro de 2025, substituindo o Voluntary AI Safety Standard de 2024). Não há legislação vinculante específica para IA, e o governo preferiu reforçar as leis existentes (privacidade, consumidor, setoriais)[78].

Implicações estratégicas

Os estudos concordam[79] que a divergência entre esses modelos (UE mais prescritiva, EUA fragmentado e setorial, China altamente centralizada, Reino Unido e Austrália mais baseados em princípios) força as empresas globais a segmentar produtos, modelos e processos de compliance por jurisdição.

Isso se traduz[80] em arquiteturas de IA de vários níveis (governança, dados, MLOps, documentação) projetadas para mapear e conciliar simultaneamente requisitos divergentes, aumentando a complexidade operacional de maneiras sem precedentes para muitos setores tradicionalmente menos regulados.

73 Patel (2025).

74 Cambridge (2025).
75 Governo do Reino Unido (2023).
76 União Interparlamentar (2025).
77 Covington (2025).
78 Governo australiano (2025).
79 Oxford (2025).
80 Banco Mundial (2024).

IA e segurança cibernética

Uma batalha com novas regras e novas superfícies de ataque

A IA está transformando radicalmente a segurança cibernética em três dimensões simultâneas: amplia os recursos ofensivos dos invasores, aprimora as defesas das organizações e, ao mesmo tempo, introduz novas vulnerabilidades que exigem proteção direcionada.

IA ofensiva: automação e adaptação em escala industrial

Os atacantes adotaram a IA com uma velocidade surpreendente, transformando métodos tradicionais em ameaças qualitativamente diferentes. O impacto é quantificável[81]: mais de 28 milhões de ataques cibernéticos com IA foram registrados globalmente em 2025, um aumento de 47% em relação ao ano anterior; o setor financeiro foi o mais afetado, com 33% desses ataques; e 87% das organizações sofreram pelo menos um ataque assistido por IA nos últimos 12 meses[82].

***Phishing* hiperpersonalizado em grande escala.** Os ataques de *phishing* gerados por IA aumentaram em 1.265% no ano desde o lançamento do ChatGPT[83], e mais de 80% dos e-mails de *phishing* agora usam modelos de linguagem para geração de texto[84]. A diferença qualitativa é dramática: enquanto o *phishing* tradicional atinge taxas de sucesso de 12%, as campanhas geradas por IA atingem taxas de cliques de 54%[85]. Os LLMs permitem a criação de mensagens personalizadas por meio da análise de perfis públicos, estilo de redação corporativa e contextos específicos da vítima, em velocidades e volumes impossíveis de serem criados manualmente.

Malware polimórfico adaptável. 76% dos malwares detectados em 2025 exibem características polimórficas alimentadas por IA[86]. Diferentemente do malware polimórfico tradicional (que sofre mutação por meio de ofuscação ou criptografia de rotina), o malware gerado por IA reescreve dinamicamente seu código em tempo real, mantendo a funcionalidade idêntica, mas com assinaturas completamente diferentes. Algumas variantes avançadas geram versões exclusivas a cada 15 segundos durante um ataque. Isso derrota os sistemas de detecção baseados em assinaturas estáticas, que historicamente têm sido a base dos antivírus tradicionais.

Mais preocupante: a IA não apenas gera variantes, mas adapta o comportamento. Os modelos de *machine learning* incorporados no malware analisam o ambiente de execução, detectam sistemas de monitoramento e ajustam suas táticas em tempo real para evitar a detecção.

***Deepfakes* e manipulação de identidade.** De acordo com a análise de segurança cibernética[87], baseada no IC3 2025, os ataques de BEC (Business E-mail Compromise) assistidos por IA aumentaram cerca de 37%, combinando texto, áudio e vídeo sintéticos para se passar por executivos. O caso mais notório: um *deepfake* de áudio do Ministro da Defesa da Itália que causou perdas financeiras significativas[88]. Em uma pesquisa, 85% das organizações relataram ter sofrido um ataque de deepfake em 2025[89].

Dark LLMs e ferramentas ofensivas especializadas. Modelos de linguagem modificados especificamente para crimes cibernéticos proliferaram: HackerGPT, WormGPT, GhostGPT, FraudGPT. Esses sistemas, criados por *jailbreaking* de modelos éticos ou modificando modelos de código aberto, são comercializados em fóruns da *dark web* com modelos de assinatura e suporte técnico. Eles geram *scripts* maliciosos, exploits e campanhas de engenharia social sem restrições éticas.

IA defensiva: detecção comportamental e resposta automatizada

As organizações estão respondendo com IA defensiva igualmente sofisticada. 51% das empresas usam IA ou automação na segurança atualmente, e a adoção está se acelerando rapidamente com evidências de ROI demonstrável[90].

Análise comportamental e detecção de anomalias. Os sistemas de análise de comportamento de usuários e instituições (UEBA) com tecnologia de IA estabelecem linhas de base dinâmicas de comportamento normal para usuários, dispositivos e aplicativos, analisando bilhões de eventos diários. Em vez de procurar por assinaturas conhecidas, eles detectam desvios sutis dos padrões estabelecidos. Esse recurso é essencial contra malware polimórfico e ataques de dia zero: em ambientes de alto risco, os sistemas baseados em IA atingem taxas de detecção de até 98%[91], diante de ameaças conhecidas ou que apresentam padrões anômalos reconhecíveis. Diante de ataques genuinamente inéditos — sem assinatura nem padrão comportamental prévio — a detecção baseada em comportamento reduz o risco, mas não elimina a incerteza: a IA defensiva não reconhece a nova ameaça, mas detecta seu desvio da norma, o que implica que ataques suficientemente cautelosos ou bem projetados para imitar um comportamento legítimo podem escapar desta detecção inicial.

Plataformas SIEM/XDR/SOAR com IA integrada. As plataformas atuais de gerenciamento de informações e eventos de segurança (SIEM), detecção e resposta estendidas (XDR) e orquestração, automação e resposta de segurança (SOAR) integram nativamente a IA para correlacionar eventos entre sistemas diferentes, reduzir falsos positivos (até 95% de redução em implantações maduras) e automatizar a resposta. A CrowdStrike relata[92] que sua plataforma Falcon analisa 4,7 bilhões de eventos diariamente com caça a ameaças alimentada por IA 24 horas por dia, 7 dias por semana. O Microsoft Sentinel demonstrou[93] reduções de 30% no tempo médio de resposta (MTTR) por meio de correlação e análise comportamental baseadas em IA.

Impacto econômico demonstrável. De acordo com a IBM[94], as organizações que usam IA e automação extensivamente na segurança reduzem os custos médios de violação em US$ 1,9

81 Revista SQ (2025).
82 SoSaft (2025).
83 PR Newswire (2023).
84 KnowBe4 (2025).
85 AICerts (2026).
86 WatchGuard (2025).
87 The Network Installers (2025).
88 Phishcare (2025).
89 Ironscales (2025).
90 IBM (2025).
91 Proofpoint (2025).
92 CrowdStrike (2025).
93 Microsoft (2026).
94 IBM (2025).

milhão (mais de 50% menos do que as organizações sem IA) e encurtam os ciclos de contenção em 80 dias, em média. As organizações com plataformas orientadas por IA detectam ameaças 60% mais rápido[95] e alcançam 95% de precisão na detecção.

Multiplicação de capacidades. A IA atua como um multiplicador de recursos para as equipes de segurança: ela automatiza a triagem de alertas (as organizações enfrentam uma média de 4.500 alertas por dia[96]), executa manuais de resposta automática para ameaças conhecidas e permite que analistas juniores operem em níveis mais altos de eficácia. Noventa e cinco por cento dos profissionais de segurança relatam que a IA melhora sua velocidade e eficiência na prevenção, detecção, resposta e recuperação de ataques[97].

Assimetria e o dilema estratégico

Apesar dos recursos defensivos avançados, persiste uma assimetria preocupante: atualmente, a maioria das empresas não tem maturidade suficiente para combater as ameaças avançadas com IA. 78% dos CISOs afirmam que as ameaças baseadas em IA agora têm um "impacto significativo" em suas organizações[98].

A segurança cibernética se tornou uma batalha de IA contra IA, com ambos os lados operando em velocidades de máquina. Os atacantes automatizam o reconhecimento, geram explorações personalizadas e adaptam suas táticas em tempo real. Os defensores correlacionam terabytes de telemetria, preveem vetores de ataque e executam a contenção autônoma. O diferencial competitivo não está mais em ter IA, mas na sofisticação dos modelos, na qualidade dos dados de treinamento, na velocidade das atualizações de inteligência contra ameaças e na capacidade de integração entre as superfícies de ataque.

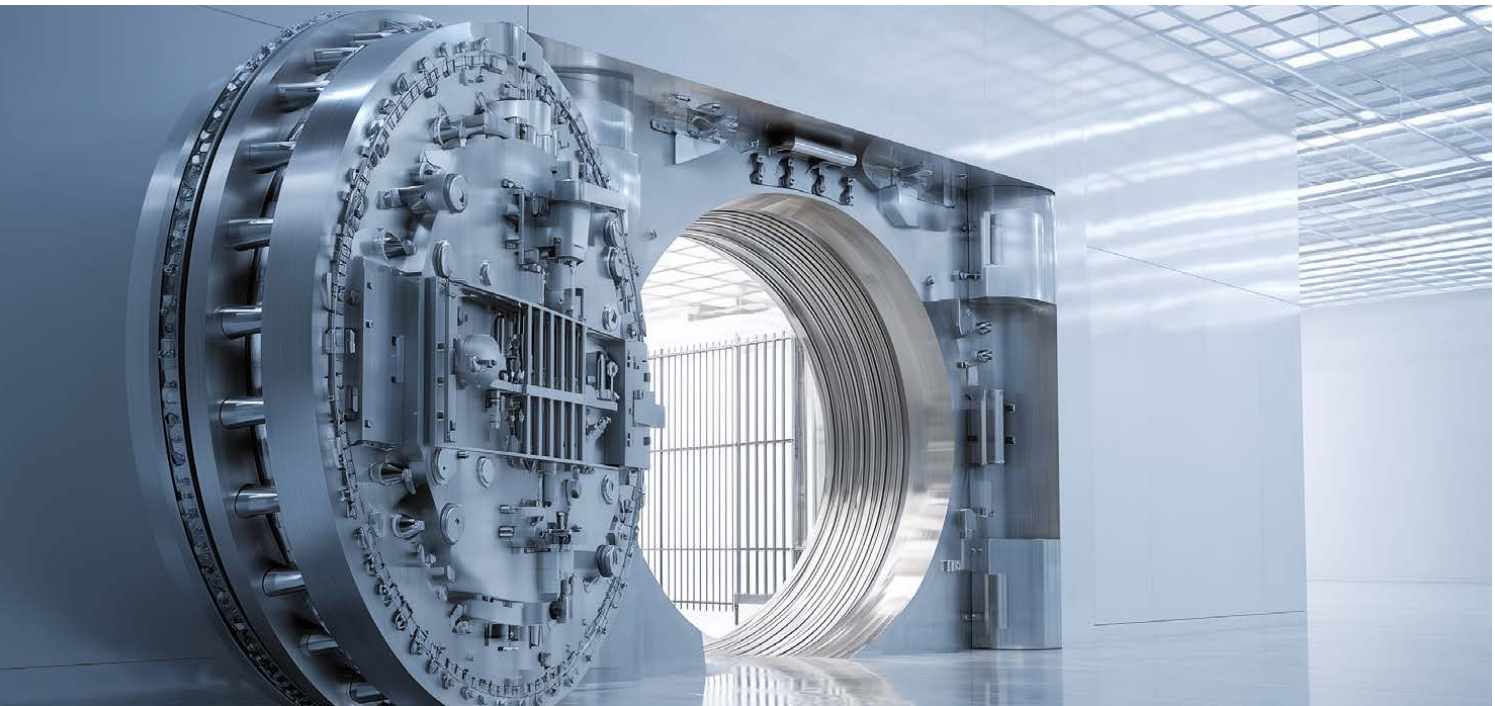
Protegendo a IA: vulnerabilidades próprias

Além da batalha entre IA ofensiva e defensiva, surge uma terceira dimensão crítica: a securitização dos próprios sistemas de IA, que introduzem vetores de ataque sem precedentes no software tradicional. Os modelos de IA são vulneráveis a ataques adversários específicos: data poisoning, em que os invasores injetam dados maliciosos para degradar o modelo; evasão adversária por meio de perturbações imperceptíveis que enganam o modelo durante a inferência; e mineração de modelos por meio de consultas repetidas para roubar propriedade intelectual[99].

Os LLMs acrescentam vetores adicionais documentados pela OWASP[100]: *prompt* injection (instruções maliciosas incorporadas em entradas que alteram o comportamento do modelo), insecure output handling (aplicativos que dependem cegamente de saídas sem validação), training data poisoning, model denial of service e vulnerabilidades da cadeia de suprimentos. O NIST publicou em 2024 diretrizes específicas para o desenvolvimento seguro de IA generativa[101], ampliando seu SSDF com controles diferenciados para cada fase do ciclo de vida. A securitização eficaz exige uma curadoria rigorosa do dataset, adversarial robustness testing, validação de inputs/outputs, sandboxing e red teaming específico de ML. As equipes de segurança devem incorporar expertise em ataques de adversários; os frameworks de governança devem considerar a IA como uma superfície de ataque independente com seus próprios controles[102].

O crime cibernético global movimenta trilhões anualmente. As organizações devem manter uma governança robusta: os próprios sistemas defensivos de IA agora são alvos de ataques (envenenamento de modelos, *prompt* injection, evasão adversária), criando uma meta camada de risco que exige proteção especializada.

95 Jumpcloud (2025).
96 HelpNetSecurity (2025).
97 Darktrace (2025).
98 Darktrace (2025).
99 NIST (2024a).
100 OWASP (2025).
101 NIST (2024a).
102 MITRE (2025).



IA, privacidade e propriedade intelectual

Um conflito estrutural

A adoção em massa da IA generativa reabre debates fundamentais sobre privacidade e propriedade intelectual, confrontando os modelos de negócios, de sistemas que exigem dados massivos, com estruturas legais projetadas para minimização e controle individual. As tensões não são meramente técnicas: elas refletem choques estruturais entre a lógica operacional dos LLMs e os princípios que regem a proteção de dados e os direitos autorais.

Privacidade em LLMs: riscos sistêmicos em todo o ciclo de vida

Em abril de 2025, o European Data Protection Board (EDPB) publicou[103] um relatório abrangente sobre os riscos de privacidade em LLMs, desenvolvido no âmbito de seu programa Support Pool of Experts. O documento identifica que cada fase do ciclo de vida do LLM introduz riscos específicos de privacidade e proteção de dados:

- **Memorização não intencional e vazamento de dados pessoais.** Os LLMs podem memorizar fragmentos de dados pessoais presentes em conjuntos de dados de treinamento e reproduzi-los posteriormente nos resultados gerados. Esse fenômeno não é um bug ocasional, mas um comportamento intrínseco: o modelo armazena padrões estatísticos que incluem dados confidenciais. O EDPB documenta casos em que *prompt*s específicos conseguiram extrair informações pessoais (nomes, e-mails, números de telefone, dados médicos) que estavam nos dados de treinamento. A magnitude do problema aumenta de acordo com o tamanho do modelo e a sensibilidade dos dados processados.
- **Reidentificação e criação de perfis não intencionais.** Embora os dados sejam anônimos antes do treinamento, as técnicas de inferência podem reidentificar indivíduos combinando várias saídas do modelo. O EDPB alerta que os LLMs podem gerar perfis detalhados de indivíduos sem o processamento explícito de dados pessoalmente identificáveis, o que viola os princípios de minimização e finalidade do GDPR.
- **Feedback loops sem salvaguardas.** As interações dos usuários com chatbots são frequentemente armazenadas para posterior ajuste fino dos modelos, de modo que os dados confidenciais revelados nas conversas são incorporados ao modelo sem consentimento explícito ou garantias de exclusão posterior.
- **Incompatibilidade estrutural com o GDPR.** O relatório do EDPB destaca tensões irresolúveis com os princípios do GDPR:
  - **Minimização de dados**: os LLMs exigem conjuntos de dados enormes, o que está em contradição direta com a minimização.
  - **Direito de ser esquecido**: ainda não existem métodos robustos para "destreinar" seletivamente os modelos (as técnicas emergentes, como o machine unlearning, ainda estão em fase experimental).
  - **Transparência**: as arquiteturas de transformadores são caixas pretas em que o rastreamento da origem de saídas específicas é tecnicamente complexo.
  - **Consentimento**: dados extraídos da Internet raramente incluem consentimento para uso em treinamento de IA.

103 EDPB (2025).

- **Avaliação de impacto obrigatória**. O EDPB conclui que, dada a natureza sistêmica do processamento nos LLMs, a realização da Data Protection Impact Assessment (DPIA) de acordo com o Art. 35 do GDPR não é apenas recomendada, mas obrigatória na maioria dos casos, especialmente quando os LLMs processam dados confidenciais ou tomam decisões com impacto sobre os indivíduos.
- **Mitigações técnicas com custo**. O relatório propõe medidas como differential privacy (adição de ruído estatístico para impedir a identificação), federated learning (modelos de treinamento sem centralização de dados), Retrieval-Augmented Generation (RAG, que separa o conhecimento atualizável da memória LLM estática) e retrospective logging minimization (minimização de dados nos registros do sistema). No entanto, todas essas técnicas implicam em precisão reduzida, alto custo computacional ou funcionalidade reduzida.

Propriedade intelectual: litígios em massa e colapso da infraestrutura

A World Intellectual Property Organization (OMPI) dedicou várias sessões de sua "WIPO Conversation on IP and AI" (Conversa da OMPI sobre PI e IA)[104] para analisar o impacto da IA generativa sobre os direitos autorais. As últimas sessões se concentraram especificamente na infraestrutura para gerenciamento de direitos, atribuição e compensação na era da IA generativa.

Dados de treinamento: uso justo ou violação em massa? O debate central não está resolvido. As empresas de IA argumentam que os modelos de treinamento com conteúdo protegido constituem "uso justo", análogo à forma como os humanos aprendem lendo. Os detentores de direitos argumentam que é a cópia não autorizada em escala industrial com intenção comercial que cria substitutos de mercado para o conteúdo original.

Os litígios estão se multiplicando; para citar alguns exemplos:

- **New York Times vs. OpenAI/Microsoft**: o NYT processa o uso de milhões de artigos sem consentimento, argumentando que os modelos criam substitutos de mercado que desviam o tráfego de seu paywall e geram alucinações que prejudicam sua reputação. Ele reivindica bilhões de dólares em danos.
- **Getty Images vs. Stability AI**: a Getty alega uso não consensual de mais de 12 milhões de fotografias para treinar o Stable Diffusion, incluindo violação de marca registrada (já que o modelo replica a marca d'água da Getty)[105].
- **Artistas vs. Midjourney/Stability AI**: os artistas alegam que suas obras foram extraídas sem permissão para treinar modelos de geração de imagens.
- **Gravadoras vs. Anthropic**: a Universal Music Group (UMG) e outras gravadoras estão processando por violação massiva do uso de letras de músicas.

Em dezembro de 2025, mais de 72 processos ativos de direitos autorais contra empresas de IA estavam em andamento[106]. Até o momento, três juízes emitiram decisões preliminares sobre o uso justo: duas decisões a favor das empresas de IA[107], uma contra[108]. As decisões finais não são esperadas até o verão de 2026, no mínimo.

Propriedade dos resultados: um vácuo jurídico - quem é o proprietário do conteúdo gerado por IA? A maioria das jurisdições (incluindo o US Copyright Office) sustenta que os direitos autorais exigem autoria humana com "input criativo suficiente". O conteúdo gerado exclusivamente por IA sem intervenção humana significativa é de domínio público. Mas os limites são imprecisos: quanta intervenção humana (*prompt* engineering, seleção, pós-edição) é "suficiente"?

Colapso da infraestrutura de direitos autorais. A WIPO alerta que os sistemas de gestão coletiva de direitos, projetados para volumes gerenciáveis de trabalhos humanos, estão entrando em colapso diante dos trilhões de resultados gerados pela IA diariamente. Não existem infraestruturas escalonáveis para rastreamento, atribuição e compensação. Sessões recentes da WIPO exploraram a necessidade de novas infraestruturas técnicas e estruturas regulatórias, mas as soluções práticas continuam especulativas.

Fragmentação e falta de consenso global

A fragmentação regulatória amplia a incerteza. A UE exige transparência nos dados de treinamento (*AI Act* Art. 53)[109], os EUA não têm legislação federal específica e a China impõe um controle estatal rigoroso sobre dados e algoritmos. As empresas que operam globalmente enfrentam requisitos não harmonizados.

As tecnologias de preservação da privacidade (differential privacy, federated learning, homomorphic encryption) oferecem caminhos técnicos para conciliar a privacidade com a utilidade da IA, mas estão longe de serem adotadas em massa: são caras, complexas e reduzem o desempenho. A tensão entre a inovação tecnológica acelerada e as estruturas jurídicas projetadas para paradigmas anteriores continua, por enquanto, sem solução.

104 WIPO (2025).
105 A Getty não obteve reconhecimento substancial na frente de direitos autorais, mas obteve uma decisão limitada sobre violação de marca registrada.
106 Chatgptiseatingtheworld (2026).
107 Bartz v. Anthropic, Kadrey v. Meta, na Califórnia.
108 Thomson Reuters v. ROSS Intelligence, em Delaware.
109 Lei de IA (2024).

05 | Governança de IA e impacto sobre as pessoas

«Nossa tecnologia reflete nossos valores».

Fei-Fei Li[110]

A adoção acelerada da IA representa um desafio que transcende o tecnológico: como governar sistemas que evoluem mais rapidamente do que as estruturas organizacionais, como transformar as funções profissionais quando as tarefas mudam trimestralmente e como preservar o julgamento humano em decisões críticas ao mesmo tempo em que se delega a operação de rotina a sistemas autônomos.

Este bloco aborda a dimensão organizacional e humana da IA: os modelos de governança corporativa que estão surgindo para controlar a IA de ponta a ponta, as práticas operacionais avançadas que permitem industrializar sua implantação (MLOps, LLMOps), a transformação de funções e perfis profissionais que já está em andamento, a adoção setorial que transforma a IA em infraestrutura transversal, o impacto na vida diária das pessoas além do trabalho, os desafios de sustentabilidade e impacto social que ela introduz e as estruturas éticas operacionais que devem traduzir princípios abstratos em controles concretos e auditoria contínua. A questão não é se a IA transformará as organizações e as pessoas, mas se seremos capazes de governar essa transformação com rigor, rapidez e responsabilidade.

110 Fei-Fei Li (nascida em 1976), cientista da computação sino-americana, pioneira da visão computacional, codiretora do Stanford Human-Centered AI Institute (HAI). Conhecida como a "madrinha da IA", ela é uma das principais defensoras da IA centrada no ser humano.

Governança corporativa da IA

A IA transborda as estruturas tradicionais

Os modelos tradicionais de governança corporativa foram projetados para tecnologias previsíveis: sistemas que executam lógica determinística, operam dentro de limites definidos e se comportam de maneira reproduzível. A IA rompe todas essas premissas: ela toma decisões sem intervenção humana, produz resultados diferentes para a mesma entrada, opera por meio de processos internos opacos que seus próprios desenvolvedores não compreendem totalmente e depende de forma crítica de fornecedores externos cujos modelos evoluem sem controle organizacional direto.

As estruturas tradicionais de governança tecnológica são insuficientes. Os comitês de arquitetura e os ciclos anuais de aprovação são lentos demais para a velocidade da inovação da IA, não têm o expertise necessário para avaliar seus riscos específicos e não foram projetados para gerenciar a incerteza inerente aos sistemas que aprendem e mudam. A questão estratégica é como governar a IA sem diminuir a velocidade de adoção e acumular riscos incontroláveis.

Organização: funções e estruturas emergentes

As organizações estão criando funções executivas especializadas, embora com progresso heterogêneo. A figura do *Chief AI Officer* (CAIO), *Chief Data and AI Officer* (CDAIO) ou equivalente está surgindo, e estima-se[111] que 26% das grandes organizações já tenham essa função[112]. Em organizações mais maduras, o CDAIO se reporta diretamente ao CEO, com uma estrutura matricial muito forte em países e empresas; em outros, a liderança recai sobre o CTO/CIO ou o diretor de inovação.

Abaixo do nível executivo, estão surgindo funções especializadas que ainda não foram padronizadas: *AI Risk Manager*, *AI Ethics Officer*, *AI Compliance Lead*, *Responsible AI Lead.* Esses perfis geralmente se reportam a funções de segunda linha, mas suas responsabilidades, autoridade e recursos variam substancialmente entre as organizações.

A primeira linha consolida o Centro de Excelência em IA, equipes transversais que centralizam o conhecimento técnico (LLMs, arquiteturas multiagentes, frameworks), infraestrutura compartilhada (plataformas cloud, GPUs, licenças), emissão de diretrizes e padrões e consultoria interna para linhas de negócios. Operacionalmente, o modelo *hub & spokes* predomina: um Centro de Excelência central estabelece recursos transversais, enquanto as equipes descentralizadas nas linhas de negócios desenvolvem soluções específicas, reportando-se hierarquicamente aos seus superiores, mas com relatórios multifuncionais para o hub. Essa estrutura permite velocidades diferentes, dependendo da maturidade de cada linha, ao mesmo tempo em que mantém a coerência arquitetônica.

Na segunda linha, a maturidade organizacional é notavelmente menor. Enquanto a primeira linha tem estruturas consolidadas, a segunda linha apresenta alta variabilidade, e há um debate conceitual aberto sobre se o *AI Risk* constitui um risco autônomo na taxonomia corporativa ou um amplificador transversal dos riscos existentes. Pouquíssimas organizações o definem como um risco de primeiro nível (como uma decisão tática para garantir visibilidade e orçamento); mais frequentemente, ele aparece como um risco de segundo nível dentro do Risco de Modelo ou dos Riscos Não Financeiros. Outros se recusam a incluí-lo formalmente e tratam a IA como um amplificador transversal dos riscos existentes (maior risco de modelo, maior risco de fornecedor, risco de tecnologia com vulnerabilidades adicionais...), reforçando as estruturas existentes em vez de criar novas taxonomias.

Independentemente do debate taxonômico, surge uma pequena função de coordenação operacional, chamada *AI Risk* ou *AI Governance*, que orquestra as avaliações de risco por funções especializadas: o Risco de Modelo valida os modelos, o Risco de Tecnologia avalia a infraestrutura, a Segurança Cibernética analisa as vulnerabilidades específicas da IA, a Proteção de Dados verifica a privacidade, o Jurídico avalia a propriedade intelectual, o *Compliance* verifica o compliance regulatório etc.

Órgãos de governança: do comitê formal ao working group operacional

Praticamente todas as organizações sistêmicas criaram um Comitê de IA com uma composição de defesa de linha cruzada: primeira linha (inovação, analytics, dados, tecnologia...), segunda linha (risco de modelo, risco operacional, segurança cibernética, proteção de dados, vendor risk, jurídico, compliance...) e terceira linha (Auditoria como observadora). A copresidência geralmente cabe a um gerente de primeira e segunda linha.

No entanto, a governança não ocorre apenas no comitê. Por baixo, geralmente há uma estrutura informal, mas fundamental: um *AI Working Group* composto por aqueles que se reportam diretamente aos membros do comitê. Eles preparam materiais, concordam com posições e resolvem conflitos. As questões chegam ao comitê *for information* ou *for approval*, não para serem discutidas do zero. A verdadeira governança (negociação, consenso, resolução de tensões entre velocidade e controle) ocorre nessa camada pré-decisória, em que a primeira e a segunda linha criam acordos operacionais antes da sanção formal.

Arquitetura de frameworks de risco: backbone e uplifting setorial

As organizações não reinventam seus frameworks de risco do zero. Observa-se uma arquitetura de duas camadas: um backbone central específico de IA (uma política de IA sucinta, que estabeleça princípios, escopo, funções, classificação de riscos, processos de aprovação; e um procedimento de IA que descreva de forma abrangente o ciclo de vida, desde a concepção até a produção) e, acima de tudo, o *uplifiting* de frameworks específicos de IA existentes.

O uplifting consiste em complementar os frameworks existentes adicionando capítulos específicos de IA. O framework de risco do modelo incorpora validação de IA generativa, técnicas de explicabilidade (SHAP, LIME, attention mechanisms), detecção de viés e alucinação, monitoramento de drift, etc. O framework de vendor risk acrescenta o due dilligence nos fornecedores de LLMs, certificações requeridas, SLAs específicos, planos de contingência,

111 IBM (2025b).

112 Entre elas estão JPMorganChase, Santander, Société Générale, Rabobank, Scotiabank, Visa, Mastercard, AXA, Pfizer, Merck, S&P, Microsoft, LinkedIn, eBay, T-Mobile, Orange, Schneider Electric, SAP, Marks and Spencer, Boeing, Nike, Michelin, para citar algumas. Veja PixieBrix (2025).

estratégias de saída. O framework de proteção de dados inclui avaliações de impacto específicas de IA, minimização de dados em sistemas de IA, exercício dos direitos do GDPR quando o processamento envolve IA. A estrutura de compliance implementa o *AI Act*: classificação de risco, registro, documentação etc.

Essa arquitetura aproveita a infraestrutura que funciona, atribui responsabilidades claras e mantém a consistência sem fragmentação. No entanto, alguns desafios técnicos são genuinamente novos. Explicar redes neurais profundas com transformadores e trilhões de parâmetros exige técnicas especializadas que não existiam na validação de modelos tradicionais. As funções de risco estão desenvolvendo esses recursos em tempo real.

Classificação de risco e ciclo de vida

O *AI Act* Europeu estabelece uma classificação (proibido, alto risco, risco limitado, risco mínimo), que é necessária, mas insuficiente para a governança corporativa das empresas. A regulação é projetada para proteger a sociedade e os direitos fundamentais, não as organizações contra seus próprios riscos operacionais, de reputação ou financeiros.

Portanto, as organizações estão convergindo para classificações internas de risco de IA mais exigentes, integrando critérios regulatórios aos seus próprios: impacto na reputação, criticidade do processo, classificação de segurança cibernética, tier do modelo de acordo com a validação, impacto financeiro direto, maturidade do fornecedor. Um sistema pode não ser de alto risco regulatório, mas pode ser de alto risco interno se afetar processos críticos ou gerar exposição massiva à reputação (Fig. 4).

A classificação determina as consequências operacionais. Por exemplo, o alto risco implica uma análise completa de todas as funções de risco, aprovação formal no comitê, documentação completa, validação independente, monitoramento intensivo etc.; enquanto o baixo risco geralmente implica um fast track, análise leve e o encaminhamento ao comitê apenas para informação. O *fast-track* é um elemento essencial, pois resolve a tensão entre velocidade e controle: obviamente, os sistemas de IA de baixo risco são aprovados sem demora, enquanto os recursos de controle são concentrados em sistemas de IA críticos.

Inventário e rastreabilidade

As organizações são obrigadas pela regulação a manter um único registro de todos os sistemas de IA implantados, incluindo sua classificação de risco e status do ciclo de vida. Esse inventário deve ser completo, atualizado e acessível aos supervisores, auditores e funções de risco.

A maioria das organizações converge para a expansão do inventário de Risco de Modelo, sob a lógica de que os sistemas de IA exigem validação quando carregam um risco de modelo relevante. Como alternativa, algumas organizações expandem o inventário de ativos de tecnologia.

A governança corporativa da IA está amadurecendo em um ritmo acelerado. Nos próximos anos, haverá uma consolidação em direção às melhores práticas, mas a aceleração da evolução tecnológica é tal que provavelmente continuará a sobrecarregar as estruturas organizacionais atuais, que foram projetadas para paradigmas mais lentos.

Fig. 4. Exemplo de classificação interna de risco de uma empresa, além da regulação.

Critérios de risco

Critérios	Acionador	Nível
Proibido por regulação	Sistema proibido/ restrição aplicável	PROIBIDO
Alto risco por regulação	Sistema de alto risco de pela regulação aplicável	ALTO
Processo	Impacta processos críticos da empresa	ALTO
Pessoas	Decisões relevantes sobre indivíduos	ALTO
Autonomia	Age de forma autônoma em nome da empresa	ALTO
Reputação	Interação direta com partes externas	ALTO
Legal	Efeitos legais ou contratuais relevantes	ALTA
Privacidade	Dados pessoais	ALTO
Confidencialidade	Dados confidenciais da empresa ou do cliente	MÉDIO
Terceiros	Dependência crítica não substituível	MÉDIO
Restante	Sem acionadores anteriores	BAIXO

+

Mitigantes

Mitigante	Acionador
Humanos in the loop	Revisão humana de 100% dos resultados
Auditoria	Auditoria periódica/avaliação técnica
Apelação	Mecanismo de apelação acessível para pessoas sujeitas ao sistema
Anonimização	Dados pessoais/ confidenciais anonimizados /pseudonimizados

=

Risco do sistema de IA

- **Risco inerente**: o nível mais alto entre os critérios de risco.
- **Risco residual**: o nível mais alto após a aplicação de atenuantes, critério a ser atendido.

PROIBIDO | ALTO | MÉDIO | BAIXO

Industrialização da IA (MLOps, LLMOps)

Da experimentação à produção

Atualmente, o principal gargalo na adoção real da IA não é algorítmico, mas operacional. Durante anos, organizações de todos os tipos desenvolveram sistemas de IA promissores em ambientes experimentais que nunca chegaram à produção ou que, uma vez implantados, falharam quando confrontados com dados reais, perderam desempenho ao longo do tempo ou geraram custos e riscos incontroláveis. A industrialização da IA surge justamente para fechar essa lacuna entre a experimentação e o uso sustentado em ambientes de produção.

O *Machine Learning Operations* (MLOps) surge como uma resposta estruturada a esse problema. Ele pode ser entendido como "um conjunto de processos padronizados e recursos tecnológicos para criar, implantar e operacionalizar sistemas de ML de forma rápida e confiável"[113]. O MLOps articula processos e recursos técnicos para gerenciar a preparação de dados, a experimentação, o treinamento, a validação, a implantação, o monitoramento e o retreinamento contínuo de modelos de forma integrada[114]. O objetivo não é apenas acelerar a produção, mas garantir a confiabilidade, a reprodutibilidade, o controle de riscos e a sustentabilidade operacional ao longo do tempo.

O advento da IA generativa amplia significativamente esse desafio. O Large Language Model Operations (LLMOps) não substitui o MLOps, mas o amplia para lidar com sistemas com propriedades radicalmente diferentes[115]. Os modelos de linguagem de grande porte introduzem comportamento não determinístico, dependência crítica da formulação imediata, arquiteturas internas opacas com bilhões ou trilhões de parâmetros e novos vetores de risco, como alucinações, geração de conteúdo inverídico ou ataques direcionados por meio da manipulação de entrada. Essas complexidades são intensificadas em sistemas agênticos, em que uma única interação do usuário pode desencadear cadeias de raciocínio, várias chamadas de modelos internos e execução autônoma de ferramentas externas.

Arquitetura do ciclo de vida de MLOps/LLMOps

A industrialização eficaz da IA exige uma visão holística do ciclo de vida operacional, que diferentes autores expressam de diferentes maneiras (Fig. 5), mas com elementos comuns:

Na fase de **preparação de dados**, o MLOps se concentra na criação de pipelines robustos para ingestão, limpeza, transformação e controle de versão de dados estruturados, garantindo rastreabilidade e qualidade. O LLMOps amplia esse escopo ao trabalhar com grandes volumes de dados não estruturados (texto, código, documentos ou imagens) e representações vetoriais usadas em arquiteturas de retrieval-oriented generation (RAG). Isso está associado a requisitos rigorosos de minimização de dados pessoais, controle de origem e compliance com estruturas de privacidade, como o GDPR.

Durante a **experimentação e o desenvolvimento**, o MLOps fornece mecanismos para rastrear experimentos, comparar modelos, gerenciar hiperparâmetros e garantir a reprodutibilidade dos resultados. No LLMOps, essa fase incorpora novas dimensões: gerenciamento de *prompt*s versionados, avaliação de diferentes configurações de modelos fundamentais, ajuste fino ou adaptação por meio de técnicas como LoRA e desenho de métricas que capturam aspectos qualitativos da geração, como consistência, factualidade, segurança ou adequação contextual. Essas métricas não substituem as métricas quantitativas tradicionais, mas as complementam quando o desempenho não pode mais ser medido apenas pela precisão ou pelo erro.

A **validação** é um dos principais pontos de divergência entre as duas abordagens. Enquanto no MLOps a validação se baseia em testes automatizados em conjuntos de dados de referência para avaliar a precisão, o viés e a robustez, no LLMOps é essencial integrar a validação humana ao ciclo. A avaliação de resultados generativos requer revisão por especialistas, técnicas de fact-*checking* em fontes confiáveis, testes de estresse semântico e exercícios de *red-teaming* projetados para identificar comportamentos indesejáveis ou vulnerabilidades exploráveis. A validação não é mais um evento isolado, mas um processo contínuo.

Na fase de **implantação**, o MLOps conta com pipelines de CI/CD relativamente maduros, com versões bem definidas de modelos e dependências. O LLMOps, por outro lado, deve lidar com implantações de modelos em grande escala com implicações significativas de infraestrutura, latência e custo. Isso inclui a seleção dinâmica de modelos com base no caso de uso, no nível de risco ou no orçamento, bem como a orquestração de sistemas complexos em que vários modelos e agentes interagem entre si de forma coordenada.

O **monitoramento** é fundamental em ambos os paradigmas, mas com ênfases diferentes. O MLOps concentra-se na detecção de *drift* de dados ou de conceitos, degradação do desempenho, erros e problemas de latência. O LLMOps acrescenta a necessidade de monitorar os custos por token em tempo real (um determinante importante da viabilidade econômica), identificar padrões de uso não previstos (*prompt drift*) e manter a rastreabilidade total das interações para análise forense, auditoria e supervisão regulatória. Sem essa visibilidade, os sistemas generativos podem aumentar rapidamente a complexidade e o custo sem um controle eficaz.

Por fim, a **governança e o compliance regulatório** abrangem todo o ciclo de vida. No MLOps, isso envolve a documentação da linhagem de dados e modelos, a definição de responsabilidades claras e a garantia de controles de qualidade. No LLMOps, esses requisitos são estendidos para responder às estruturas regulatórias emergentes, como o *AI Act*, incluindo inventários de sistemas classificados por nível de risco, mecanismos de supervisão humana e estratégias de explicabilidade adaptadas a modelos que funcionam como caixas pretas.

113 Google (2023).
114 iDanae (3T20).
115 Shan (2024).

Integração com frameworks de governança

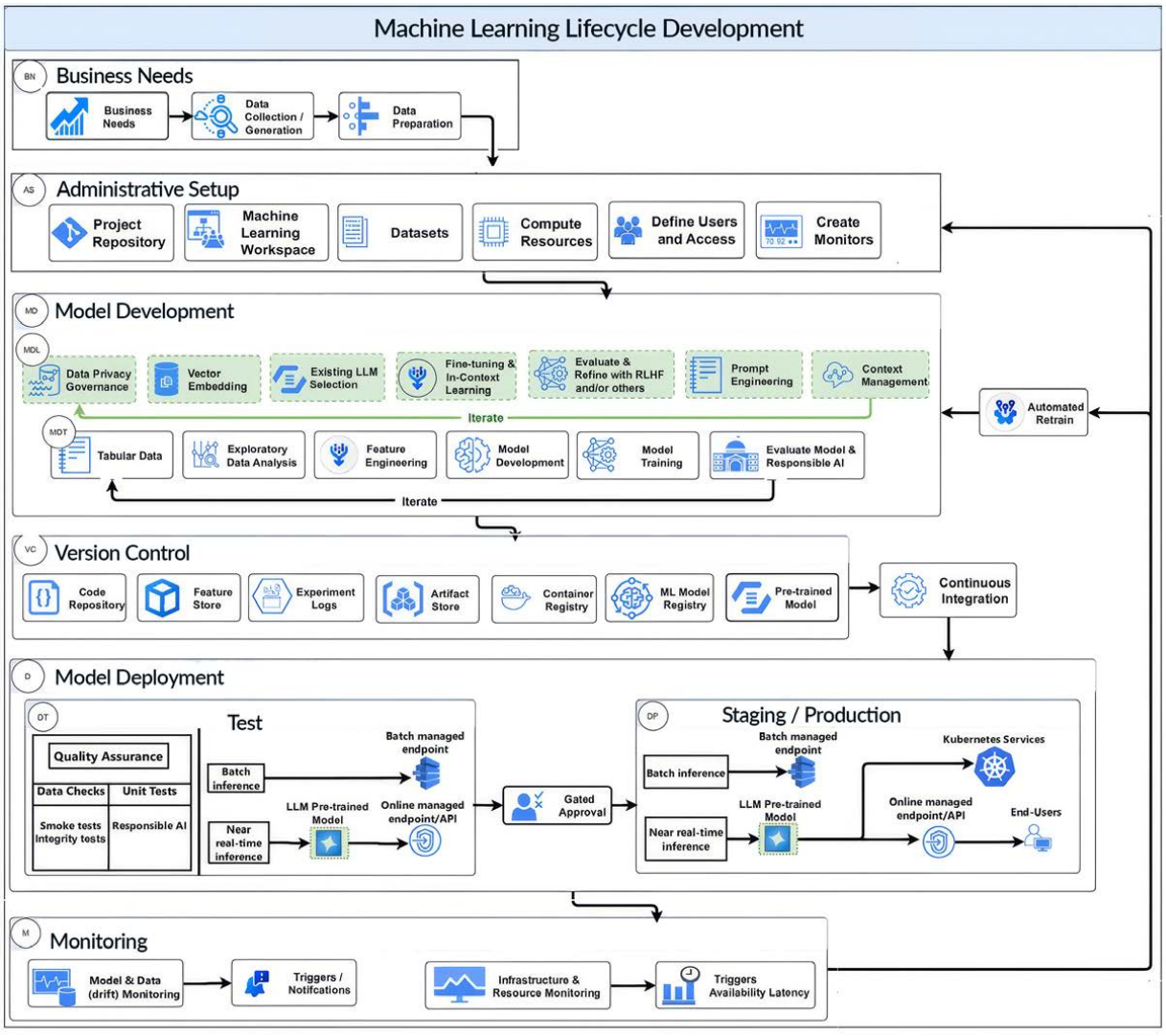
Os MLOps e LLMOps não são disciplinas puramente técnicas, nem podem operar isoladamente. Eles são a camada operacional que incorpora os princípios definidos em frameworks de governança mais amplas. O framework de gestão de riscos de IA do NIST[116] estrutura a gestão de riscos de IA como um processo contínuo que abrange governança, contextualização, mensuração e gestão ao longo de todo o ciclo de vida do sistema. Complementarmente, a ISO/IEC 42001 define[117] os requisitos de um sistema de gestão de IA, incluindo políticas, avaliações de impacto, controle de fornecedores e monitoramento contínuo.

Os processos automatizados de MLOps e LLMOps (pipelines de dados, validações sistemáticas, implementações controladas e monitoramento contínuo) são os mecanismos concretos que permitem que estes frameworks de governança passem da teoria à prática. Sem uma base operacional sólida, a governança de IA é reduzida a uma documentação aspiracional; sem frameworks de governança claros, a industrialização técnica carece de critérios de qualidade, responsabilidade e controle de riscos.

Portanto, a adoção madura da IA exige uma convergência real entre engenharia, operações e governança. A industrialização da IA não se trata apenas de modelos de escala, mas da criação de sistemas confiáveis, auditáveis e sustentáveis que possam ser integrados de forma responsável aos processos comerciais essenciais e à vida cotidiana das organizações. MLOps e LLMOps são frameworks de boas práticas em constante evolução, não soluções fechadas: se o problema de levar a IA à produção estivesse resolvido, não se observaria a persistência de modelos que se deterioram silenciosamente, geram custos imprevistos ou falham ao serem escalonados. Seu valor está em reduzir o atrito e estruturar a gestão de riscos, não em eliminá-los..

116 NIST (2023).
117 ISO/IEC (2023).

Fig. 5. Exemplo de fases de MLOps e LLMOps. Fonte: Stone (2025).



Upskilling, reskilling e novas funções profissionais

O fator humano, o gargalo da IA

A proliferação da IA está transformando o trabalho de maneiras mais profundas do que sugerem as discussões focadas apenas na automação ou substituição de tarefas. O principal desafio para as organizações não é tecnológico, mas humano: ter as habilidades certas para desenhar, implantar, operar e governar sistemas de IA de forma sustentável. Nesse contexto, o talento não é mais entendido como um conjunto restrito de especialistas, mas como uma competência organizacional distribuída.

As conclusões desta seção são apoiadas por uma análise empírica abrangente realizada pela Management Solutions em um conjunto de 16 grandes organizações na Europa e nos EUA, com base em dados reais de mercado e evidências organizacionais.

Convergência em direção a um núcleo estável de funções de IA

À medida em que a IA amadurece, as organizações estão convergindo para um conjunto relativamente estável de competências e capacidades profissionais. Embora a nomenclatura varie entre os setores e as empresas, as funções desempenhadas por esses perfis estão se tornando homogêneas. No entanto, as funções formais são menos importantes do que as competências subjacentes: na prática, um mesmo profissional combina várias dessas competências, e as combinações mais incomuns (quem domina ao mesmo tempo LLMOps, regulação e prompt engineering, por exemplo) são precisamente as mais valiosas e escassas. Essa convergência reflete uma realidade operacional: o ciclo de vida da IA – dos dados à produção e ao controle – exige recursos diferenciados que não podem ser concentrados em um único tipo de profissional.

Em resumo, essas funções podem ser agrupadas em três blocos principais (Fig. 6). Primeiro, os perfis técnicos, responsáveis por desenhar modelos, construir infraestruturas de dados e levar soluções à produção. Em segundo lugar, os perfis híbridos, que conectam os recursos técnicos às necessidades de negócio e garantem que os sistemas de IA gerem valor real e sejam adotados. Por fim, os perfis de governança e controle, responsáveis pela gestão de riscos, compliance, auditoria e uso responsável da IA.

Essa estrutura de competências forma a espinha dorsal sobre a qual os recursos de IA das organizações são construídos, independentemente do setor em que operam.

Maturidade desigual na adoção de funções especializadas

Embora o núcleo básico das funções de IA seja amplamente implantado em organizações avançadas (com heterogeneidade significativa), a adoção de perfis emergentes ou mais especializados é muito desigual em termos de maturidade. As diferenças não estão tanto na existência de recursos gerais, mas na institucionalização de funções avançadas relacionadas à operação de IA generativa e agêntica, à arquitetura de larga escala ou à governança específica desses sistemas.

Essa heterogeneidade é especialmente visível em funções emergentes ligadas a LLMOps, avaliação contínua de modelos generativos, segurança de *prompt*s ou governança de IA. Em alguns casos, essas funções existem como papéis formais; em outros, elas são informalmente integradas a equipes mais tradicionais, com resultados mistos em termos de controle, escalabilidade e custo.

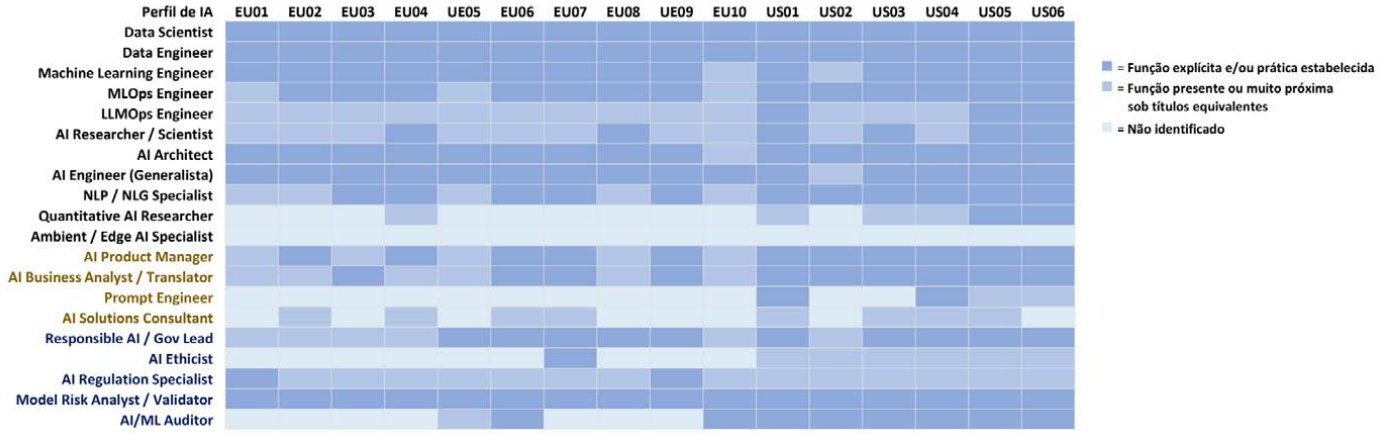
O mapa de calor[118] (Fig. 7) ilustra que a diferença entre as organizações se manifesta em quais funções elas consolidaram e com que nível de especialização e autonomia.

118 O mapa de calor é baseado em uma análise comparativa da adoção de funções de IA em um conjunto de 16 grandes organizações europeias e norte-americanas. As funções foram identificadas a partir de fontes públicas (vagas de emprego, estruturas organizacionais, iniciativas estratégicas anunciadas) e evidências qualitativas do mercado, considerando tanto as funções explícitas quanto as funções equivalentes com nomes alternativos. A intensidade da cor reflete o grau de institucionalização da função, desde a ausência do perfil até a prática consolidada.

Fig. 6: Perfis consolidados e emergentes de IA.

#	Perfil de IA	Tipo de perfil	Descrição
01	Data Scientist	Técnico	Projetar modelos preditivos e extrair insights acionáveis para melhorar os produtos e as decisões do banco.
02	Data Engineer	Técnico	Cria e mantém a infraestrutura e os pipelines de dados que alimentam modelos e análises.
03	Machine Learning Engineer	Técnico	Implanta modelos na produção e desenvolve APIs/serviços para a empresa consumir com segurança e eficiência.
04	MLOps Engineer	Técnico	Automatiza e opera o ciclo de vida dos modelos (CI/CD, monitoramento, retraining), garantindo a robustez.
05	LLMOps Engineer	Técnico	Industrializa o GenAI na produção (routing, caching, observabilidade, avaliação contínua, guardrails, implantação e custo).
06	AI Researcher / Scientist	Técnico	Pesquisa novas técnicas de IA/ML e desenvolve protótipos avançados que podem ser convertidos em recursos de bancada.
07	AI Architect	Técnico	Projeta a arquitetura das soluções de IA, integrando modelos aos sistemas core e garantindo escalabilidade e segurança.
08	AI Engineer (Generalist)	Técnico	Cria soluções de software que integram componentes de IA pré-treinados em aplicativos bancários.
09	NLP / NLG Specialist	Técnico	Desenvolve modelos de linguagem e soluções baseadas em LLM para analisar, resumir e gerar conteúdo textual.
10	Quantitative AI Researcher	Técnico	Aplica IA e deep learning a mercados e estratégias quantitativas para melhorar os sinais, a previsão e o trading.
11	Ambient / Edge AI Specialist	Técnico	Implementa IA em ambientes IoT/edge e sistemas de ambiente (sensores, dispositivos), integrando dados em tempo real.
12	AI Product Manager	Híbrido	Lidera o ciclo de vida do produto de IA, traduz os recursos técnicos em valor comercial e prioriza sua evolução.
13	AI Business Analyst / Translator	Híbrido	Identifica casos de uso de IA e conecta as necessidades de negócios com requisitos técnicos acionáveis.
14	Prompt Engineer	Híbrido	Projeta, testa e padroniza prompts para casos de uso de GenAI e protege os aplicativos GenAI contra prompt injection e jailbreaks.
15	AI Solutions Consultant	Híbrido	Lidera a implementação de soluções de IA em áreas de negócios, garantindo valor real e adoção efetiva.
16	Responsible AI / AI Governance Lead	Governança	Define políticas responsáveis de IA e supervisiona modelos de conformidade ética, regulatória e de risco.
17	AI Ethicist	Governança	Define e operacionaliza critérios éticos (viés, impacto social...) em casos de uso de IA, com guardrails e revisão de decisões.
18	AI Regulation Specialist	Governança	Traduz regulação (AI Act e análogos), políticas internas e expectativas supervisoras em requisitos, controles, evidências e reporting.
19	Model Risk Analyst / Validator	Governança	Valida modelos de IA/ML por meio de análise independente para garantir robustez, imparcialidade e compliance regulatório.
20	AI/ML Auditor	Governança	Audita processos, controles e uso de AI para verificar a conformidade com políticas internas e regulações.

Fig. 7. Mapa de calor da adoção de funções de IA.



O verdadeiro gargalo: desequilíbrios entre oferta e demanda

A análise de oferta e demanda[119] (Fig. 8) mostra um desequilíbrio estrutural generalizado no mercado de talentos de IA. A demanda é alta em praticamente todos os perfis analisados, enquanto a oferta é sistematicamente insuficiente para absorvê-la. Não se trata de uma escassez pontual ou concentrada, mas de um fenômeno transversal que afeta a maior parte do ecossistema profissional de IA.

A única exceção parcial é o perfil do cientista de dados, que está mais próximo de um equilíbrio relativo. Esse comportamento se deve à sua maior maturidade histórica, à existência de caminhos de treinamento consolidados e a um pool maior de profissionais com experiência acumulada. Entretanto, mesmo nesse caso, a pressão da demanda continua alta em cargos sênior ou especialistas.

119 A estimativa da demanda baseia-se em uma análise longitudinal das ofertas de emprego publicadas entre o segundo semestre de 2025 e janeiro de 2026 por grandes organizações europeias e norte-americanas, considerando o volume relativo por perfil, a senioridade exigida e a recorrência temporal. Essa análise é complementada por relatórios setoriais do mercado de trabalho e evidências qualitativas de iniciativas estratégicas de IA (criação de centros de excelência, funções de IA responsáveis, risco de modelo e GenAI).

A oferta é estimada em termos relativos com base na dificuldade de cobertura observada no mercado, incluindo prêmios salariais, senioridade necessária e escassez relatada em estudos especializados. Além disso, a maturidade histórica de cada perfil é incorporada (senioridade da disciplina, existência de pipelines de treinamento) e a viabilidade de transição de funções adjacentes por meio de upskilling ou reskilling, em contraste com as observações do mercado sobre a disponibilidade efetiva de profissionais com experiência demonstrável.

No restante dos perfis, especialmente aqueles voltados para a produção (*Machine Learning* Engineering, MLOps, LLMOps) e funções de governança, risco e controle, o gap entre a demanda e a oferta é persistente. A combinação de complexidade técnica, requisitos de senioridade e aumento das exigências regulatórias limita a capacidade do mercado de gerar talentos no ritmo necessário. Como resultado, a terceirização, por si só, não permite fechar o gap, fazendo com que o upskilling e o reskilling internos sejam uma alavanca estrutural para escalar a IA com impacto e controle.

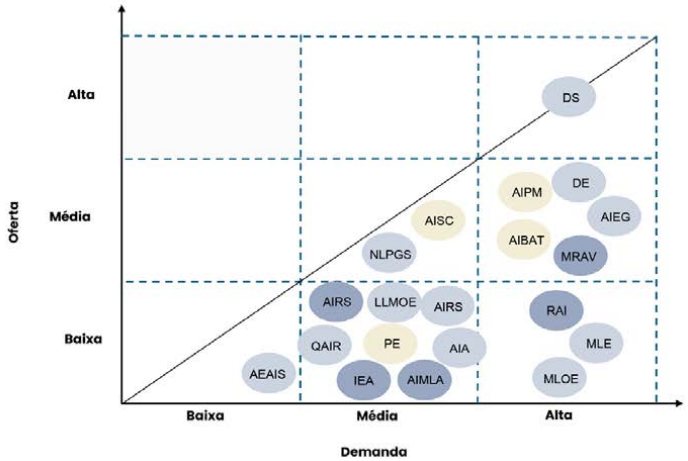
Implicações estratégicas

As implicações dessa análise são claras. A estratégia de talentos de IA não pode se basear exclusivamente na atração de perfis escassos no mercado. O aprimoramento e a requalificação internos tornam-se alavancas estratégicas, especialmente para funções críticas em que a oferta externa é estruturalmente insuficiente.

Da mesma forma, os perfis operacionais e de governança são tão decisivos quanto os perfis de modelagem ou pesquisa. A vantagem competitiva não está apenas no desenvolvimento de modelos avançados, mas na integração deles de forma segura, eficiente e compatível com os processos reais da organização.

Em última análise, a transformação impulsionada pela IA não é apenas tecnológica. É uma transformação do trabalho, das funções profissionais e das capacidades coletivas. As organizações que entenderem essa dimensão humana e a abordarem de forma estruturada estarão mais bem posicionadas para capturar o valor da IA de forma sustentável.

Fig. 8. Matriz de oferta e demanda de perfis de IA



Acrônimo	Perfil IA
DS	Data Scientist
DE	Data Engineer
MLE	Machine Learning Engineer
MLOE	MLOps Engineer
LLMOE	LLMOps Engineer
AIRS	AI Researcher / Scientist
AIA	AI Architect
AIEG	AI Engineer (Generalist)
NLPGS	NLP / NLG Specialist
QAIR	Quantitative AI Researcher
AEAIS	Ambient / Edge AI Specialist
AIPM	AI Product Manager
AIBAT	AI Business Analyst / Translator
PE	Prompt Engineer
AISC	AI Solutions Consultant
RAI	Responsible AI / AI Governance Lead
AIE	AI Ethicist
AIRS	AI Regulation Specialist
MRAV	Model Risk Analyst / Validator
AIMLA	AI/ML Auditor

IA e transformação do setor (AI + X)

A IA como uma camada transversal

A IA deixou de ser uma tecnologia aplicada de forma incremental em setores específicos para se tornar uma camada transversal de inteligência que é integrada simultaneamente em vários domínios econômicos e sociais. Essa transição, amplamente documentada por organizações internacionais, é caracterizada por uma mudança de pilotos isolados para uma adoção estrutural que afeta processos inteiros, cadeias de valor e modelos operacionais. Nem todo o avanço segue uma lógica transversal: alguns dos desenvolvimentos mais significativos são impulsionados por setores específicos (IA médica, defesa ou serviços financeiros regulados), nos quais a especificidade do domínio, a natureza dos dados ou o marco regulatório geram capacidades que dificilmente podem ser transferidas para outros contextos sem uma adaptação substancial.

Em uma perspectiva comparativa, as evidências mostram que as diferenças entre os setores não são mais explicadas pela mera presença da IA, mas pela intensidade e profundidade de sua integração. A OCDE[120] classifica os setores de acordo com sua *AI Intensity* com base em fatores como composição de tarefas, disponibilidade de dados e capital humano, mostrando que mesmo os setores tradicionalmente pouco digitalizados estão aumentando rapidamente sua exposição à IA. Essa adoção ocorre em paralelo entre os setores, gerando efeitos de aceleração cruzada e aprendizado intersetorial[121].

Em termos macroeconômicos e de emprego, o impacto é sistêmico. O Fundo Monetário Internacional estima[122] que cerca de 40% do emprego global esteja exposto à IA, com porcentagens mais altas nas economias avançadas, onde a complementaridade entre IA e mão de obra qualificada é maior. A Organização Internacional do Trabalho qualifica[123] que a IA generativa tende a automatizar tarefas específicas em vez de ocupações inteiras, reforçando a necessidade de adaptação organizacional e treinamento contínuo.

Nesse contexto, o conceito AI + X é útil para descrever um padrão estrutural: a IA é combinada com domínios setoriais específicos sem perder seu caráter de tecnologia de uso geral. Não se trata de uma soma de casos de uso independentes, mas de uma integração simultânea e transversal, apoiada por evidências empíricas recentes[124.] Esta seção apresenta alguns exemplos ilustrativos, sem a pretensão de ser exaustiva (Fig. 9).

Saúde (AI + Health): precisão clínica e escalabilidade

Na área da saúde, a IA está sendo integrada a diagnósticos clínicos, suporte a decisões médicas, monitoramento de pacientes e automação administrativa. A Organização Mundial da Saúde documenta[125] um crescimento constante no uso de sistemas de IA em radiologia, patologia e atenção primária, destacando tanto seu potencial clínico quanto os riscos associados.

Vários estudos[126] demonstram que os sistemas de IA atingem níveis de precisão comparáveis ou superiores aos de especialistas humanos em tarefas como detecção de câncer em imagens médicas ou classificação dermatológica. Esses resultados não implicam em substituição direta, mas em uma expansão significativa da capacidade de diagnóstico e triagem em larga escala. Esse desenvolvimento, enfatiza a OMS, exige estruturas de governança sólidas, validação clínica e supervisão humana, devido ao seu impacto direto sobre as pessoas.

Educação (AI + Education): personalização em escala

Na educação, a IA generativa está transformando a criação de conteúdo, a avaliação e o suporte ao aprendizado. A UNESCO identifica[127] o uso crescente de tutores adaptativos, a geração automática de materiais de aprendizagem e os sistemas de avaliação formativa contínua como tendo o potencial de melhorar a personalização e reduzir as gaps educacionais.

A Educação[128] é um dos setores em que a IA pode ter impactos estruturais no médio prazo, permitindo modelos de aprendizagem mais flexíveis e adaptáveis. No entanto, esses benefícios dependem fundamentalmente de políticas públicas e estruturas institucionais que preservem a equidade, a integridade acadêmica e o papel dos professores.

Trabalho e emprego (AI + Work): reconfiguração de tarefas

A IA está redefinindo o trabalho de forma transversal, afetando as ocupações em praticamente todos os setores. O FMI estima[129] que a exposição à IA é responsável por cerca de 60% do emprego nas economias avançadas, em comparação com porcentagens menores nas economias emergentes, refletindo as diferenças na estrutura de produção e no capital humano.

A OIT, por sua vez, conclui[130] que a IA generativa tem, por enquanto, um impacto maior em tarefas cognitivas específicas do que em empregos inteiros, o que implica uma reconfiguração do conteúdo do emprego em vez de uma substituição massiva. Essa evidência reforça a necessidade de estratégias de aprimoramento e requalificação, conforme discutido acima.

Indústria e operações (AI + Industry): produtividade e eficiência

Em ambientes industriais e operacionais, a IA é aplicada à manutenção preditiva, à otimização de processos, ao planejamento e à robótica avançada. Muitas organizações[131] já passaram de testes-piloto para implantações em escala total, integrando a IA a processos críticos de produção e logística.

Já existem fortes evidências[132] de melhorias significativas de eficiência e redução de falhas em sistemas industriais em que a IA é sistematicamente integrada, embora os maiores benefícios sejam observados quando a tecnologia é acompanhada por mudanças organizacionais e de processo.

120 OECD (2024).
121 Fórum Econômico Mundial (2025a).
122 FMI (2024).
123 OIT (2025a).
124 Stanford (2025).
125 OMS (2024).
126 Stanford (2025).
127 UNESCO (2023).
128 Fórum Econômico Mundial (2025b).
129 FMI (2024).
130 OIT (2025a).
131 Fórum Econômico Mundial (2025a).
132 Stanford HAI (2025).

Finanças e serviços intensivos em informação (AI + Finance)

Em setores com uso intensivo de informações, como finanças e serviços profissionais, a IA é usada para detecção de fraudes, gestão de riscos, personalização de serviços e automação de documentos. Uma expansão constante do uso da IA nessas áreas é observada[133], com diferenças relevantes dependendo da estrutura regulatória e da capacidade organizacional.

A adoção da IA generativa nesses setores está associada a melhorias de produtividade e ao surgimento de novos modelos de serviços, embora seja observado[134] que os benefícios dependem da qualidade dos dados, da governança e da integração aos processos existentes.

Criatividade e conteúdo (AI + Criative): novos modelos culturais

A geração de texto, imagem, música e vídeo orientada por IA está transformando os setores criativos e de mídia. Vemos[135] um crescimento acelerado desses aplicativos e um aumento significativo em seu uso comercial e cultural nos últimos anos.

Esse desenvolvimento levanta debates regulatórios e econômicos relevantes, especialmente em relação à propriedade intelectual e aos modelos de negócios, que estão sendo abordados de forma desigual por diferentes estruturas regulatórias.

Síntese: de setores à infraestrutura transversal

Além dos exemplos setoriais, as evidências convergem para um ponto central: a transformação atual não está na adoção isolada da IA em setores específicos, mas em sua integração simultânea e transversal como infraestrutura operacional e cognitiva. Os estudos concordam que a vantagem competitiva não é mais obtida com a aplicação da IA a funções individuais, mas com sua integração coerente ao longo de toda a cadeia de valor.

Nesse sentido, a AI+X não descreve mais um modismo terminológico, mas um padrão estrutural de transformação. Compreender essa lógica é essencial para prever os impactos setoriais, elaborar políticas públicas adequadas e definir estratégias organizacionais capazes de capturar o valor da IA de forma sustentável.

133 OCDE (2025a); OCDE (2026).
134 OCDE (2025b).
135 Stanford (2025).

Fig. 9. Exemplos representativos de AI + X, priorizando casos em que a IA passou do piloto para o uso operacional mensurável. Fonte: adaptado de Stanford (2025).

Domínio (X)	Exemplo AI + X	Impacto
Saúde - diagnóstico	Sistemas de IA já detectam câncer em imagens médicas com maior precisão do que os humanos.	Permite a triagem em massa e o diagnóstico precoce em uma escala impossível para equipes humanas.
Saúde - prática clínica	Médicos usam assistentes de IA que ouvem a consulta e escrevem o histórico médico.	Dezenas de minutos por dia de burocracia médica reduzidos; mais tempo para o paciente, menos esgotamento.
Mobilidade urbana	Robotáxis operam continuamente em várias cidades, com centenas de milhares de viagens por semana.	O transporte autônomo deixa de ser um experimento e se torna um serviço urbano real.
Logística	Caminhões autônomos transportam mercadorias em grande escala, com milhões de quilômetros acumulados.	Custos reduzidos, operação 24 horas por dia, 7 dias por semana e menos dependência da falta de motoristas.
Indústria	Robôs industriais habilitados para IA são instalados em ritmo recorde, especialmente na Ásia.	A automação avançada se torna uma vantagem competitiva estrutural, e não pontual.
Robótica avançada	Modelos de IA permitem que os robôs aprendam tarefas complexas com menos programação manual.	O desenvolvimento e a implementação de robôs versáteis se aceleram drasticamente.
Engenharia de software	Programadores usam a IA para escrever, depurar e documentar códigos sistematicamente.	O desenvolvimento de software acelera e muda a função do programador, que passa a ser mais designer do que digitador.
Atendimento ao cliente	Sistemas generativos gerenciam interações completas com os clientes nos centros de contato.	Aumento da produtividade e escalação de serviços sem crescimento proporcional de pessoal.
Supply chain	A IA otimiza os estoques, a previsão de demanda e o planejamento logístico em tempo real.	Menos quebras de estoque, menos excesso de estoque, maior eficiência operacional.
Marketing e vendas	Geração automática de campanhas, conteúdo e segmentação personalizada em grande escala.	Aumento mensurável da receita e redução do tempo de comercialização.
Estratégia corporativa	Executivos usam a IA para analisar cenários, documentos estratégicos e alternativas de decisão.	O planejamento estratégico é apoiado por análises contínuas, não por exercícios anuais.
Finanças	IA automatiza a análise financeira, a detecção de anomalias e o suporte a decisões de investimento.	Maior velocidade e da cobertura analítica sem um aumento proporcional das equipes.
Jurídico	Revisão automatizada de contratos e pesquisa jurídica com IA generativa.	Os processos jurídicos que costumavam levar semanas são reduzidos a horas ou dias.
Recursos humanos	IA dá suporte ao recrutamento, à avaliação e à gestão de talentos com análise massiva de dados.	Processos mais rápidos e mais consistentes, com riscos que exigem governança explícita.
Educação	Tutores de IA adaptam o conteúdo e o ritmo ao nível de cada aluno.	Aprendizagem mais personalizada, mesmo em ambientes com alta taxa aluno-professor.
Criatividade	IA gera músicas, imagens e vídeos que são integrados a processos criativos profissionais.	Ela muda o processo criativo: da produção manual para a cocriação homem-máquina.
Mídia e comunicação	Geração automática de conteúdo de notícias, resumos e traduções.	Aumento da produção editorial com novos desafios de qualidade e precisão.
Segurança cibernética	IA detecta padrões de ataque e anomalias em tempo real.	Recursos defensivos aprimorados contra ameaças crescentes e automatizadas.
Operações internas	IA automatiza tarefas transversais (documentação, relatórios, análise interna).	Redução estrutural dos custos administrativos e do atrito organizacional.
Economia digital	A maioria das grandes organizações já usa IA generativa em pelo menos uma função crítica.	A IA deixa de ser um diferencial e se torna uma infraestrutura básica.

IA na vida pessoal e cotidiana

A mudança de paradigma: pessoas em primeiro lugar, organizações em segundo

A IA generativa foi uma inversão radical do paradigma tecnológico tradicional. Ao contrário das tecnologias empresariais anteriores (cloud computing, ERP, CRM), que nasceram em ambientes corporativos e foram gradualmente transpostas para o consumo pessoal[136], a IA generativa entrou primeiro na vida cotidiana das pessoas e só depois foi formalmente adotada pelas organizações.

Os dados são convincentes: estima-se em[137] que o ChatGPT atingirá aproximadamente 900 milhões de usuários ativos semanais até o final de 2025. Esse número coloca a IA generativa entre as tecnologias de adoção mais rápida da história, superando a taxa de penetração da Internet móvel, das redes sociais ou dos serviços de streaming.

Mais revelador ainda é o fato de que a proporção de uso pessoal não apenas precede, mas também excede consistentemente o uso profissional. Na União Europeia[138,] 25,1% da população usa ferramentas de IA generativas para fins pessoais, em comparação com 15,1% em contextos de trabalho e apenas 9,4% na educação formal. Nos países da OCDE[139,] mais de um terço das pessoas relatam o uso regular de IA generativa, sendo que os estudantes lideram a adoção: três quartos dos estudantes com mais de 16 anos usam essas ferramentas, enquanto 41,1% dos funcionários as integram em seu trabalho, muitas vezes informalmente, mesmo antes de suas organizações autorizarem.

Essa sequência temporal não é anedótica: ela redefine a lógica da transformação digital. As organizações não estão liderando a adoção da IA; elas estão respondendo a recursos que seus funcionários, clientes e fornecedores já possuem e usam de forma não oficial. O fenômeno da IA invisível (uso não autorizado de ferramentas de IA em ambientes corporativos) não é uma anomalia transitória, mas uma consequência estrutural desse investimento: as pessoas adquiriram habilidades cognitivas ampliadas em suas vidas pessoais que, em seguida, transferiram espontaneamente para seus empregos, criando exposições a riscos que a maioria das empresas ainda não controla.

Magnitude, distribuição e fraturas: quem usa IA e com que finalidade

A adoção da IA na vida pessoal abrange um amplo espectro de atividades. Os usos dominantes incluem criação de conteúdo (texto, imagens, música, vídeo), suporte para aprendizado e exploração conceitual, automação de tarefas rotineiras, companhia para conversas e acesso a informações: nos EUA, 10% dos adultos usam chatbots de IA para obter notícias[140,] introduzindo um novo vetor de mediação de informações com profundas implicações para o debate público.

Entretanto, essa democratização é radicalmente assimétrica. As divisões geográficas são extremas[141:] na Noruega, 56% da população usa ferramentas de IA generativas; na Dinamarca, 48,4%; na Estônia, 46,6%. Em contrapartida, na Turquia, o número cai para 17%; na Romênia, 17,8%. Mesmo dentro das economias desenvolvidas, as diferenças são substanciais: a Alemanha está em 32%, a França em 37%, a Espanha em 38%, enquanto a Itália permanece abaixo de 20%.

A diferença de gerações é ainda mais acentuada[142.] A diferença na adoção entre os grupos etários mais jovens e mais velhos chega a 53,6 pontos percentuais, o que a torna o fator de segmentação mais decisivo, à frente da educação (diferença de 21 pontos percentuais) ou do nível de renda (21 pontos). Enquanto 75% dos estudantes usam IA generativa regularmente, apenas 12,5% dos aposentados ou pessoas economicamente inativas relatam tê-la usado. A

136 De forma diferente da Internet, que primeiro se tornou de uso geral e somente depois corporativo.
137 Backlinko (2025).
138 Eurostat (2025).
139 OCDE (2026).
140 Pew (2025).
141 Eurostat (2025).
142 OCDE (2026).

diferença de gênero, por outro lado, é comparativamente menor: 4,2 pontos percentuais.

O paradoxo da adoção em massa

Globalmente, 66% da população acredita[143] que os produtos e serviços baseados em IA terão um impacto significativo em suas vidas diárias nos próximos três a cinco anos, um aumento de seis pontos percentuais desde 2022. No entanto, esse reconhecimento do impacto iminente coexiste com uma ambivalência crescente.

O público em geral demonstra um alto nível de preocupação[144:] nos EUA, 51% dos adultos dizem que estão mais preocupados do que animados com o aumento do uso da IA na vida cotidiana, em comparação com apenas 11% que expressam maior entusiasmo. Entre os especialistas em IA, a proporção é inversa: 47% estão mais animados do que preocupados e apenas 15% mais preocupados. Essa divergência de 40 pontos percentuais entre os especialistas e o público reflete não apenas diferenças no conhecimento técnico, mas também percepções estruturalmente diferentes da relação risco-benefício.

A familiaridade com a IA não leva automaticamente à confiança. Embora o uso pessoal se correlacione positivamente com as percepções da IA como uma oportunidade[145,] ele também aumenta a conscientização de seus riscos. A confiança de que as empresas de IA protegem adequadamente os dados pessoais varia de 50% em 2023 a 47% em 2024[146.] Ao mesmo tempo, diminuiu a proporção de pessoas que acreditam que os sistemas de IA são imparciais e livres de discriminação.

O contexto de uso é absolutamente fundamental. Pesquisas no Reino Unido mostram oscilações de até 110 pontos percentuais na aceitação pública, dependendo do caso de uso específico: de um saldo líquido de +53% de conforto com o uso da IA para analisar dados de tráfego em tempo real e melhorar os fluxos rodoviários, a um saldo de -57% contra o uso da IA para analisar preferências políticas e direcionar publicidade personalizada[147].

Há, no entanto, um consenso transversal[148] entre os especialistas e o público em geral: ambos os grupos querem mais controle sobre como a IA é usada em suas vidas (55% do público e 57% dos especialistas), e ambos acham que não têm esse controle atualmente (menos de 25% em ambos os casos). Este gap entre a demanda por atuação e a percepção de impotência constitui um dos desafios mais urgentes para a governança democrática da IA.

Desigualdade de acesso e risco de exclusão

A assimetria na adoção da IA na vida pessoal não é uma exclusão digital convencional. Não se trata apenas de uma questão de acesso à infraestrutura tecnológica (conexão à Internet, dispositivos), mas de acesso diferenciado a uma maior capacidade intelectual. Aqueles que integram a IA como uma ferramenta cognitiva cotidiana obtêm vantagens cumulativas em termos de aprendizado, produtividade, criatividade e acesso à informação. Aqueles que não o fazem enfrentam um atraso estrutural crescente.

Os grupos desconectados digitalmente percebem a IA de forma marcadamente negativa: 51% preveem um impacto pessoal negativo, em comparação com apenas 17% que esperam benefícios[149]. Essa percepção não é irracional: ela reflete a intuição de que a transformação em andamento pode redistribuir as oportunidades de forma profundamente desigual.

A IA já está transformando a vida diária de aproximadamente um bilhão de pessoas. A questão estratégica não é mais se essa transformação continuará, mas se as sociedades criarão estruturas que transformem a IA em uma infraestrutura pública inclusiva ou se permitirão que ela se enraíze como um vetor de fragmentação social irreversível.

143 Stanford (2025).
144 Pew (2025).
145 DSIT DO REINO UNIDO (2024).
146 Stanford (2025).
147 DSIT DO REINO UNIDO (2024).
148 Pew (2025).
149 UK DSIT (2024).

IA, sustentabilidade e impacto social

A sustentabilidade entra na era algorítmica

A transição sustentável não é mais uma questão exclusiva de infraestruturas físicas (novas usinas renováveis, redes de eletricidade, eletrificação do transporte), mas também um desafio de coordenação sistêmica. Os sistemas de energia atuais combinam geração distribuída, intermitência renovável, armazenamento, mercados dinâmicos e demanda flexível. Essa complexidade não pode ser gerenciada apenas por regras estáticas ou planejamento linear.

Nesse contexto, a IA surge como uma infraestrutura cognitiva capaz de modelar interdependências, simular cenários e otimizar decisões em tempo real. A eletrificação acelerada e o crescimento estrutural da demanda de eletricidade estão remodelando a capacidade, a flexibilidade e as necessidades de planejamento da rede nos próximos anos: em 2024, os data centers consumiram 1,5% da eletricidade global[150], a energia necessária para treinar modelos de fronteira está crescendo a uma taxa de 2,2x a 2,9x por ano (Fig. 10), e as maiores execuções já excedem 100 MW de energia instantânea[151]. Portanto, a sustentabilidade está entrando em uma fase em que a capacidade computacional se torna parte integrante da arquitetura de energia e clima.

A mudança é conceitual: de decisões baseadas em médias históricas para decisões baseadas em simulações dinâmicas e análises probabilísticas de alta resolução. A IA não substitui a engenharia e a física, mas amplia a capacidade de antecipação e coordenação em sistemas com múltiplas variáveis e restrições simultâneas.

Otimização do planeta: eficiência, resiliência e adaptação

Em termos operacionais, a IA está agindo como um acelerador da transição. Nas redes de eletricidade, os modelos preditivos permitem antecipar picos de demanda, otimizar o envio, reduzir o congestionamento e melhorar a integração de energias renováveis variáveis. Sem a IA, a integração ideal das energias renováveis e o gerenciamento em tempo real seriam mais incertos. Em um ambiente em que a eletricidade assume um papel central na descarbonização, a eficiência operacional deixa de ser marginal e se torna uma condição estrutural.

A relação entre energia e IA também é bidirecional: embora a IA possa melhorar significativamente a eficiência energética, o planejamento da rede e o gerenciamento de sistemas complexos e reduzir as emissões da ordem de 1.400 Mt CO_2eq por ano até 2035 em cenários de ampla adoção[152], No entanto, este potencial de redução coexiste com um aumento estrutural do consumo energético da própria infraestrutura de IA que, na ausência de uma transição paralela para energias limpas, pode resultar em um saldo líquido neutro ou negativo em termos de emissões. A questão estrategicamente relevante não é se a IA pode contribuir para a descarbonização —ela pode—, mas sob quais condições energéticas e de governança este potencial se materializa sem ser consumido por sua própria pegada infraestrutural.

Além do sistema elétrico, a IA melhora a resiliência a riscos físicos crescentes. A modelagem climática apoiada no *machine learning* aumenta a granularidade espacial e temporal das previsões, facilitando as decisões em planejamento urbano, seguros e infraestrutura essencial. Na agricultura e no gerenciamento de água, a otimização orientada por dados reduz o desperdício e ajusta os insumos de acordo com as restrições ambientais.

Esses desenvolvimentos estão inseridos em uma agenda global mais ampla. O progresso em direção aos Objetivos de Desenvolvimento Sustentável está progredindo[153] em um cenário de estresse energético, climático e demográfico. Nesse cenário, a IA pode atuar como um catalisador transversal: não como um substituto para as políticas públicas, mas como uma ferramenta que melhora a eficiência, reduz os atritos e permite alocações mais precisas de recursos escassos.

O custo oculto: energia, materiais e concentração de infraestrutura

A própria infraestrutura digital que possibilita essas melhorias gera novas pressões. O crescimento dos data centers e das cargas computacionais associadas à IA está contribuindo para o aumento da demanda de eletricidade em algumas economias avançadas:

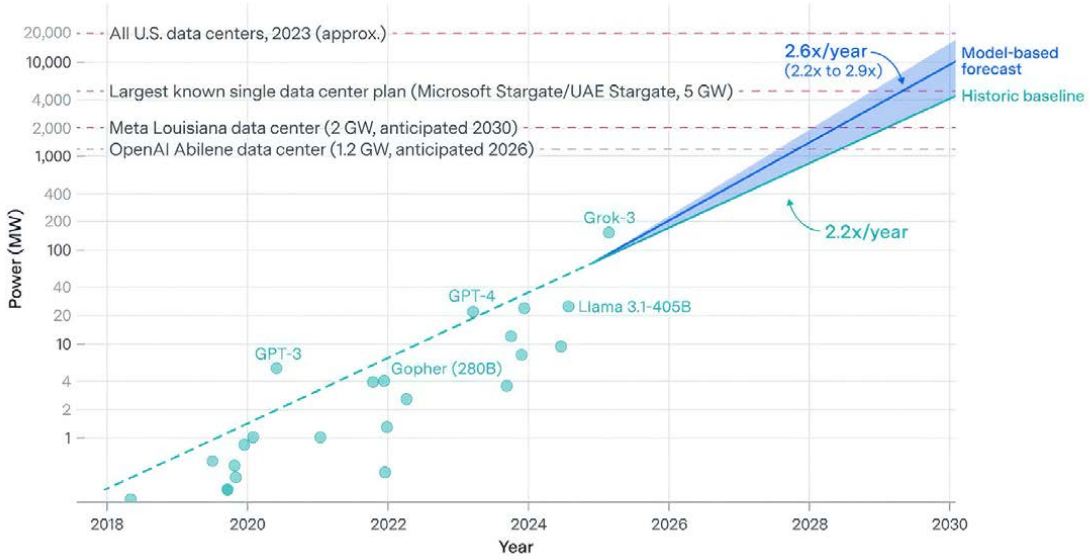
- A AIE projeta que o consumo global de eletricidade dos data centers se multiplicará até 2030, atingindo 945 TWh/ano, o equivalente ao consumo de eletricidade do Japão atualmente, sendo a IA o principal impulsionador dessa expansão[154].
- As maiores execuções individuais de treinamento de modelos de fronteira podem exigir de 4 a 16 GW de energia até 2030[155], o equivalente à geração de várias usinas nucleares e ao consumo de eletricidade de milhões de residências[156].
- A energia necessária para treinar modelos de fronteira tem aumentado mais de 2 vezes por ano, impulsionada pelo crescimento da computação de 4 a 5 vezes por ano, parcialmente compensado por melhorias na eficiência energética do hardware (cerca de 40% por ano nas principais GPUs)[157].

Portanto, a implantação massiva de modelos avançados pode se tornar um impulsionador estrutural do crescimento do consumo de energia se não for acompanhada por melhorias de eficiência e planejamento energético[158].

A sustentabilidade da IA não pode ser avaliada apenas pelos benefícios que produz, mas também pelos recursos que consome: em cenários de linha de base, as emissões de CO_2 da eletricidade do data center podem chegar a 300-320 Mt de CO_2 por ano até 2030 se a eletricidade adicional continuar a depender muito de combustíveis fósseis[159]. O treinamento e a implantação de modelos em larga escala requerem energia, água para resfriamento (em algumas regiões, em concorrência direta com o uso agrícola ou

150 IEA (2025a).
151 Epoch (2025b).
152 IEA (2025b).
153 Nações Unidas (2025).
154 IEA (2025a).
155 Estas projeções não ignoram os avanços em eficiência: a Nvidia prevê uma melhoria de 10 vezes no desempenho por watt ao passar da arquitetura Blackwell para a Vera Rubin, e as melhorias na eficiência energética das principais GPUs giram em torno de 40% ao ano. O fato de os números de consumo projetados serem o que são, apesar destes ganhos, reflete a magnitude do crescimento na demanda computacional que os neutraliza.
156 Epoch (2025b).
157 Epoch (2025b).
158 IEA (2025b).
159 IEA (2025a).

Fig. 10. Crescimento da demanda de energia da IA de fronteira. Fonte: Epoch (2025b)..



doméstico) e hardware com uso intensivo de materiais críticos. Além disso, a localização geográfica dos data centers introduz assimetrias: a intensidade de carbono da eletricidade varia substancialmente entre as regiões, o que implica pegadas ambientais diferenciadas para o mesmo serviço digital[160].

Essa dimensão de infraestrutura acrescenta uma camada geopolítica. A concentração da capacidade de computação em determinados países ou regiões reconfigura as dependências estratégicas e o acesso à tecnologia. A IA não é apenas uma ferramenta ambiental; ela também é um nó crítico no mapa global de energia e tecnologia.

A questão não é se a IA "compensa" sua pegada, mas como ela é projetada e implantada para maximizar a eficiência energética por unidade de capacidade cognitiva gerada. A sustentabilidade da IA torna-se, portanto, uma questão de arquitetura e de governança.

Sustentabilidade sem inclusão não é sustentável

A dimensão ambiental não esgota a análise. Deve-se fazer uma distinção entre a transição energética socialmente justa (ligada à mudança no modelo de produção e energia) e a transição justa associada à implantação da própria IA. Essa última introduz efeitos distributivos específicos. A capacidade de integrar tecnologias avançadas depende do capital humano, da infraestrutura digital e da qualidade institucional. As economias mais intensivas em tecnologia tendem a capturar os ganhos de produtividade e eficiência mais cedo, o que pode aumentar as diferenças com as regiões menos preparadas[161].

Do ponto de vista do desenvolvimento humano, a tecnologia pode expandir as capacidades (educação, acesso a serviços, participação econômica) ou consolidar as exclusões existentes, dependendo do contexto de adoção e do grau de democratização do acesso às suas ferramentas[162]. A ampliação de capacidades digitais em PMEs, empresários e grupos com menos capital tecnológico constitui um mecanismo importante para dar suporte tanto à criação quanto ao ajuste. Na esfera do trabalho, a convergência entre a transição verde e a automação inteligente pode acelerar os processos de realocação setorial e transformação de habilidades, com uma sequência de tempo em que os efeitos de deslocamento podem preceder a criação de novas oportunidades. O gerenciamento desse intervalo de tempo exige planejamento estratégico, políticas ativas de qualificação e mecanismos para facilitar a ampla disseminação dos ganhos de produtividade.

De uma perspectiva estrutural, a sustentabilidade ambiental exige uma transição energética socialmente justa que evite a concentração dos custos da mudança em determinados territórios, setores ou grupos. Esse desafio é conceitualmente independente do uso da IA. Paralelamente, a implantação da IA apresenta sua própria transição justa: os ganhos de produtividade e eficiência tendem a se materializar de forma desigual e com uma defasagem de tempo em relação aos impactos negativos iniciais, especialmente no emprego e em determinadas habilidades. Portanto, a IA aplicada à sustentabilidade deve ser avaliada não apenas em termos de seu impacto agregado, mas também em termos da distribuição efetiva de seus benefícios e das estratégias públicas e privadas para ampliar o acesso e atenuar os custos sociais do ajuste.

IA sustentável by design: da ambição à disciplina

A convergência dessas dimensões – otimização sistêmica, pressão infraestrutural e efeitos distributivos – força o debate a mudar de princípios gerais para práticas verificáveis. A IA alinhada às metas de sustentabilidade exige métricas explícitas sobre o consumo de energia e as emissões associadas, transparência na arquitetura e no local de implantação e avaliação dos impactos sociais nas decisões automatizadas.

A estrutura dos ODS reforça essa necessidade de coerência estrutural entre tecnologia e desenvolvimento sustentável[163]. Por sua vez, a inter-relação entre energia e digitalização exige a integração de variáveis energéticas no planejamento tecnológico corporativo e público.

Em suma, a IA ocupa uma posição ambivalente na transição sustentável: ela pode possibilitar reduções de emissões da ordem de gigatoneladas[164], mas o treinamento de seus modelos mais avançados pode atingir escalas de 4 a 16 GW por execução individual até 2030, colocando a IA na mesma escala de energia das grandes infraestruturas industriais[165].

A IA é simultaneamente um acelerador da eficiência sistêmica, um novo ônus de infraestrutura e uma força distributiva com efeitos sociais diferenciados. A questão não é se a IA fará parte da transição sustentável, mas sob quais condições de energia e distribuição.

160 iDanae (1T24).
161 Banco Mundial (2025).
162 PNUD (2025).
163 Nações Unidas (2025).
164 IEA (2025b).
165 Epoch (2025b).

Ética e filosofia da IA

Um problema em aberto após mais de 200 frameworks

A OIT catalogou 245 frameworks, códigos e recomendações de ética de IA emitidos desde 2017 por governos, agências internacionais, empresas e sociedade civil[166], uma meta-análise acadêmica anterior[167] identificou pelo menos 17 princípios recorrentes em 200 deles: transparência, justiça, responsabilidade, privacidade, há um consenso regulatório subjacente. E, no entanto, os problemas éticos associados à IA não diminuíram com a proliferação destes frameworks, mas aumentaram em complexidade, e surgiram questões para as quais nenhum deles oferecia categorias anteriores.

O gap entre a formulação de princípios e sua tradução em decisões concretas não é um problema técnico a ser resolvido: é o problema central da governança da IA.

Do princípio à ação: operacionalizando a ética da IA

O gap entre um documento de valores e o comportamento real de um sistema de IA é onde reside o risco operacional. Esse fenômeno, conhecido como *ethics washing*, nem sempre é deliberado: ele geralmente reflete a dificuldade genuína de traduzir um princípio abstrato ("o sistema deve ser justo") em um critério de design, um procedimento de auditoria ou uma linha organizacional de responsabilidade. Para resolver este gap, é necessário transformar a abordagem da ética de declarativa para operacional.

Um modelo de operacionalização bem documentado e amplamente citado no setor financeiro é o De Volksbank, um banco holandês cuja governança ética de IA foi analisada em detalhes[168]. Sua estrutura inclui uma área de Ética de IA incorporada à função de Compliance, com perfis filosóficos e técnicos, que avalia cada sistema de IA individualmente antes da implantação: análise de impacto ético, triagem de viés, rastreabilidade de decisões e canais de escalonamento formalizados. O caso ilustra que a operacionalização eficaz não está no escopo dos princípios declarados, mas na robustez dos processos, funções e pontos de decisão que os concretizam.

Resumindo o estado da arte[169], uma estrutura operacional de ética de IA em uma grande organização normalmente incorpora seis componentes:

1. **Estrutura de governança**: definição de quem decide, quem revisa e quais linhas de defesa existem, com documentação explícita das responsabilidades.
2. **Avaliação do impacto no sistema**: análise específica para cada caso de uso, proporcional ao nível de autonomia do sistema e às consequências das decisões delegadas a ele.
3. **Gestão contínua de vieses**: mecanismos de detecção e correção que operam de forma contínua, não como auditorias pontuais.

166 OIT (2025b).
167 Corrêa (2023).
168 Krijger (2023).
169 NIST (2023).

4. **Transparência e explicabilidade diferenciada por público**: os requisitos de informação do órgão regulador, do cliente e do funcionário afetado por uma decisão automatizada são substancialmente diferentes.
5. **Mecanismos de escalação e comunicação**: canais acessíveis e protegidos para comunicar comportamentos inesperados do sistema.
6. **Revisão periódica do próprio framework**: as atualizações dos modelos podem alterar o perfil de risco dos usos já implementados, exigindo reavaliação contínua, não apenas no momento da implementação inicial.

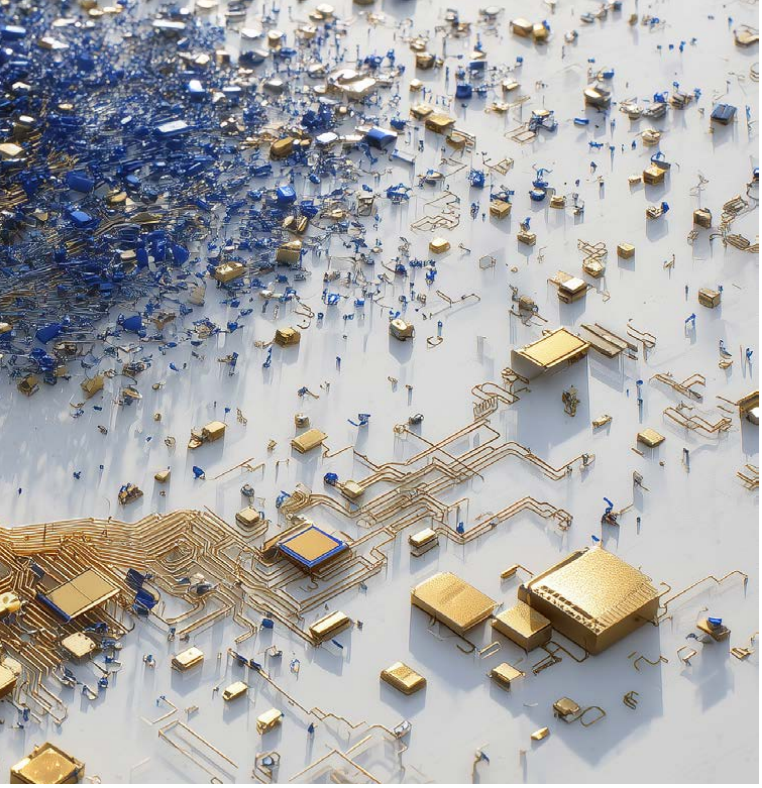
Um benchmark de operacionalização recente atua em um nível mais profundo: não na camada de implantação organizacional, mas no treinamento do próprio modelo. A Constitution publicada pela Anthropic[170] em janeiro de 2026 é o primeiro documento público de um laboratório de fronteira a codificar valores diretamente no processo de treinamento, com uma hierarquia explícita e fundamentada entre segurança, ética, compliance e utilidade. A mudança conceitual subjacente é significativa: em vez de impor regras de comportamento de fora, o sistema pretende internalizar o raciocínio por trás de cada princípio, de modo que possa generalizar esse julgamento para situações imprevistas.

A mudança da função ética para a operacionalização enfrenta, no entanto, um limite conceitual que os frameworks atuais não resolvem: eles pressupõem implicitamente que há clareza sobre o tipo de instituição que está sendo governada.

O que estamos criando? A pergunta que as estruturas não respondem

As estruturas regulatórias existentes, incluindo o *AI Act* da UE[171], classificam os sistemas de IA por nível de risco e domínio de aplicação. Essa classificação é operacionalmente útil e oferece alguma cobertura ética (no sentido de proteger contra possíveis abusos da IA), mas não faz distinção entre os diferentes tipos de relacionamento que um sistema estabelece com os seres humanos. Um sistema de scoring de crédito, um assistente de conversação e um agente autônomo que negocia contratos em nome de uma organização podem se sobrepor em sua categoria de risco regulatório e, ainda assim, operar em bases éticas

170 Anthropic (2026).
171 Lei da IA (2024).



fundamentalmente diferentes. Um sistema que adapta seu comportamento ao interlocutor, mantém a consistência ao longo do tempo e produz respostas contextualmente indistinguíveis das de uma pessoa com compreensão genuína levanta questões que o compliance regulatório convencional não foi projetado para absorver.

Essas questões são de quatro tipos, todas acionáveis para as organizações que estão implantando sistemas de IA:

- **Ontologia**: que tipo de instituição é esse sistema e quais categorias são relevantes para descrevê-lo e governá-lo?
- **Epistemologia**: como o sistema verifica o que afirma e como os humanos podem validar a confiabilidade do que ele produz?
- **Teoria da mente**: existe algum tipo de experiência interna associada à operação do sistema (por mais rudimentar que seja) e que implicações isso tem para aqueles que o projetam e implantam?
- **Ética aplicada**: que obrigações o sistema cria para a organização que o utiliza, além do que é explicitamente exigido pela regulação?

Essas não são perguntas especulativas. Em janeiro de 2026, a Anthropic reconheceu publicamente[172] que seu modelo Claude "pode possuir alguma forma de consciência ou status moral", tornando-se o primeiro laboratório de fronteira a tornar pública essa afirmação. A importância desse reconhecimento não está apenas no que ele diz, mas no que ele revela: que uma empresa de classe mundial admite que não pode responder com certeza à pergunta sobre o que ela criou.

Todas as decisões sobre como usar, auditar e regular um sistema de IA incorporam implicitamente uma resposta a essa pergunta. Na maioria das organizações, essa resposta ocorre por padrão, sem deliberação explícita.

Quatro fraturas: perguntas sem resposta

A expansão da IA não apenas amplia os dilemas éticos conhecidos: ela gera fraturas conceituais para as quais as estruturas legais e éticas atuais não têm uma resposta estruturada.

A primeira é a crise **epistêmica da verificação**. O AlphaFold previu a estrutura de mais de 200 milhões de proteínas, e mais de três milhões de pesquisadores em 190 países as utilizam como base de seu trabalho[173]. Uma proporção significativa dessas estruturas não pode ser verificada experimentalmente com métodos científicos convencionais, mas elas são usadas para projetar medicamentos e orientar decisões clínicas. A questão subjacente não é se o sistema funciona (ele o faz com um nível de precisão sem precedentes), mas quais protocolos éticos e regulatórios correspondem a um cenário em que o conhecimento cientificamente operacional se torna computacionalmente inacessível à verificação humana.

A segunda é a **assimetria de responsabilidade em danos emergentes**. As estruturas éticas e jurídicas herdadas pressupõem atores identificáveis com intenções discerníveis. Os sistemas de IA causam danos sem intenção deliberada direta e com a causalidade distribuída entre vários atores (projetistas, treinadores, implantadores e usuários), configurando o que a literatura[174] chama de "*many hands problem*": situações em que a responsabilidade moral e legal é estruturalmente diluída à medida que a cadeia causal se fragmenta. As estruturas de responsabilidade e os regimes de aplicação atuais não têm uma resposta estruturada para esse tipo de causalidade difusa.

A terceira é a **representação de futuros não nascidos**. Os sistemas de IA são treinados com base em dados históricos. Seus vieses são os do passado humano, amplificados em escala. Atualmente, nenhuma estrutura ética de IA incorpora mecanismos[175] que representem os interesses de gerações que ainda não nasceram e, portanto, não produziram dados, mas que viverão com as consequências dos sistemas desenhados no presente. As implicações são diretas para aplicações com um longo escopo temporal, desde o planejamento urbano até a modelagem de riscos climáticos.

O quarto é o **livre arbítrio como um risco sistêmico**. Uma proporção cada vez maior de decisões individuais e organizacionais é de fato mediada por sistemas de IA cuja oferta é altamente concentrada: três fornecedores respondem por 88% dos gastos globais das empresas com modelos de linguagem[176], e o mercado de infraestrutura de nuvem subjacente é igualmente concentrado. Essa estrutura implica que um viés introduzido deliberada ou acidentalmente em um desses modelos não afeta decisões individuais, mas milhões de processos simultâneos em diferentes organizações, setores e países. A diversidade cognitiva de uma sociedade (ou seja, sua capacidade de chegar a conclusões diferentes por caminhos diferentes) depende, em parte, de os sistemas que mediam seu pensamento não serem homogêneos nem oligopolistas.

Essas quatro fraturas não são argumentos contra a adoção da IA, mas o território conceitual que as organizações que levam a sério o gerenciamento da adoção da IA devem aprender a habitar. A relevância dessas fraturas não é diminuída pela ausência delas nas estruturas regulatórias existentes; pelo contrário, é justamente a ausência delas que as torna os riscos com menor visibilidade e maior potencial de dano..

172 Anthropic (2026).

173 Google DeepMind (2025).
174 Thompson (1980), Nissembaum (1996).
175 Rawls (1971), Parfit (1984).
176 CEP (2026).

06 | Fronteiras da IA

«Como alguém que é muito bom em acompanhar a IA, mal consigo me manter atualizado sobre ela».

Ethan Mollick[177]

As tendências discutidas até agora descrevem transformações que já estão ocorrendo: sistemas em produção, regulamentações em vigor, organizações se adaptando. Este bloco trata de uma dimensão diferente. As seis tendências que o compõem não se limitam ao presente operacional; elas delimitam o espaço estratégico que já condiciona as decisões de investimento, o posicionamento e a soberania, mesmo que seus efeitos completos ainda estejam se revelando.

A geopolítica da IA redefine alianças e dependências. As organizações que priorizam a IA preveem modelos competitivos sem precedentes. Os gêmeos digitais e a simulação avançada alteram a maneira como projetamos, experimentamos e decidimos. A ambient AI obscurece a fronteira entre o ambiente e a computação. A convergência com a computação quântica amplia o horizonte de problemas abordáveis. E a AGI deixou de ser mera especulação acadêmica e se tornou uma hipótese estratégica explícita nos principais laboratórios do mundo.

177 Ethan Mollick (n. 1975), professor associado da Wharton School da Universidade da Pensilvânia e pesquisador de IA, autor do best-seller do New York Times Co-Intelligence. Seu trabalho se concentra no impacto da IA no trabalho e na educação, e ele foi reconhecido pela revista TIME como uma das pessoas mais influentes em IA em 2024.

Geopolítica e soberania tecnológica da IA

A IA como um ativo estratégico do Estado

A IA não é mais uma tecnologia setorial, mas uma infraestrutura estratégica comparável aos sistemas de energia, telecomunicações ou financeiro. O que está em jogo não é apenas a competitividade econômica: é a capacidade dos Estados de manter a autonomia nas decisões críticas, desde a defesa até a supervisão financeira e o gerenciamento da infraestrutura crítica. Nesse contexto, o controle de modelos fundamentais, semicondutores avançados, centros de dados e talentos especializados tornou-se um importante fator de poder nacional, e as políticas industriais, os controles de exportação e as estratégias de investimento já refletem essa nova realidade.

A cadeia de valor estratégico: onde a soberania está em jogo

A soberania tecnológica em IA não é um estado binário: ela se desenvolve em camadas, e a dependência pode ser gerada em qualquer uma delas:

- **Hardware**: a TSMC fabrica mais de 90% dos chips mais avançados do mundo, a ASML é a única fornecedora global da litografia EUV necessária para produzi-los, e a NVIDIA domina, com mais de 85% de participação, o mercado de GPUs para treinamento de modelos de fronteira. Três empresas, três países, um gargalo estrutural.
- **Infraestrutura e modelos**: três hiperescaladores (AWS, Azure, Google Cloud) são responsáveis pela maior parte da capacidade global de computação, com quase dois terços do total; além disso, a OpenAI, a Anthropic e a Google DeepMind desenvolvem os modelos básicos mais capazes, com vantagens cumulativas em dados, talento e investimento que são difíceis de reproduzir.
- **Talento**: a pesquisa de fronteira permanece geograficamente concentrada, com a mobilidade internacional transformando a política de migração em política tecnológica.

Nenhum país ou organização controla todas essas camadas simultaneamente. A questão estratégica não é quantas camadas são controladas, mas quais são de missão crítica e quais são gerenciáveis por meio de diversificação, acordos ou redundância.

Três modelos concorrentes

Os Estados Unidos, a China e a Europa articularam respostas estruturalmente distintas que não são apenas diferenças de ênfase, mas projetos geopolíticos com profundas consequências para alianças, mercados e padrões globais.

Os **Estados Unidos** combinam a primazia do setor privado no desenvolvimento de modelos com a crescente intervenção estatal na cadeia de suprimentos: o CHIPS and Science Act mobiliza dezenas de bilhões em semicondutores domésticos, e os controles de exportação de chips avançados para a China representam a maior restrição tecnológica entre as principais potências desde a Guerra Fria. A estratégia é clara: manter a vantagem em modelos de ponta e privar os concorrentes do hardware necessário para alcançá-la.

A **China** combina investimento estatal massivo, integração civil-militar e uma estratégia explícita de autossuficiência tecnológica e controle abrangente do ecossistema digital. Sua meta declarada é a autossuficiência em toda a cadeia de valor até 2030, desde semicondutores até modelos básicos desenvolvidos internamente. As restrições de exportação dos EUA aceleraram essa agenda: a DeepSeek demonstrou em 2025 que a inovação chinesa pode produzir modelos competitivos com hardware da geração anterior, complicando a lógica de contenção por meio do controle de chips.

A **Europa** adotou a liderança regulatória como um vetor de influência geopolítica. O *AI Act* e o GDPR geraram um verdadeiro "efeito Bruxelas": as empresas globais adaptam seus produtos aos padrões europeus porque o mercado europeu é grande demais para ser ignorado. No entanto, a Europa mantém dependências infraestruturais significativas: seus modelos básicos mais avançados são americanos, sua nuvem é em grande parte estrangeira e sua capacidade de computação soberana é limitada. A ASML é a exceção notável: o monopólio holandês na litografia EUV faz da Europa um participante indispensável na cadeia global de semicondutores.

O restante do mundo navega entre esses três polos com capacidades muito desiguais. Alguns países emergentes estão articulando suas próprias estratégias com ambição cada vez maior: a Índia está desenvolvendo modelos fundamentais em idiomas locais e negociando sua posição nas cadeias de suprimentos de semicondutores; o Brasil está liderando iniciativas de governança de IA na América Latina; a União Africana está promovendo estruturas continentais para a soberania digital. No entanto, para a maioria dos países, a escolha entre ecossistemas incompatíveis permanece implícita e não deliberada, com riscos reais de dependência estrutural que a literatura chama de "colonialismo de dados": a mineração de dados locais para treinar modelos que são implantados globalmente, sem que os países de origem capturem valor ou mantenham o controle efetivo sobre sua infraestrutura digital.

Fragmentação e o risco de dissociação

O mundo está caminhando para ecossistemas tecnológicos parcialmente incompatíveis. Controles de exportação de chips, dados geo-fenced, padrões técnicos divergentes e infraestruturas paralelas estão moldando o que alguns analistas chamam de "technoblocks": esferas de influência tecnológica com suas próprias lógicas de governança, segurança e valores. A aliança Chip 4 (EUA, Japão, Taiwan e Coreia do Sul) coordena a estratégia ocidental de semicondutores; a China cultiva sua própria esfera por meio da Digital Silk Road. Uma dissociação completa entre os ecossistemas dos EUA e da China forçaria países terceiros e organizações a escolher, com custos de transição potencialmente proibitivos.

Os limites da soberania absoluta

A autossuficiência total em IA é economicamente inviável para a maioria dos países e organizações. Recriar toda a cadeia de valor (desde a mineração de minerais essenciais até o desenvolvimento de modelos fundamentais) do zero exige investimentos e escalas que somente dois ou três participantes globais podem sustentar. A soberania tecnológica realista não é isolamento: é a capacidade efetiva de decidir, diversificar fornecedores, negociar termos e evitar o aprisionamento estratégico nas camadas realmente críticas. Para a maioria dos países, a soberania da IA na prática se resume ao controle do único ativo sobre o qual eles têm jurisdição efetiva: os dados gerados em seu território.

Implicações para as organizações

O debate geopolítico se reflete em decisões corporativas concretas. A dependência de um único provedor de modelos básicos expõe as organizações a riscos de lock-in, mudanças nos preços ou termos de serviço e até mesmo restrições regulatórias decorrentes de tensões entre jurisdições. A localização de dados multijurisdicionais e o compliance regulatório acrescentam camadas de complexidade operacional crescente.

As estratégias de vários modelos e várias nuvens, antes justificadas por motivos de desempenho técnico e custo, agora assumem uma dimensão estratégica adicional: elas são o equivalente organizacional da diversificação de dependências soberanas.

Organizações *AI-first* e *AI-only*

Definição e taxonomia

A expansão dos sistemas agênticos levanta uma questão que, até alguns anos atrás, era teórica: uma organização pode funcionar com a IA como sua arquitetura cognitiva central, relegando o trabalho humano à exceção? Para responder a essa pergunta com precisão, é útil distinguir três estágios que costumam ser confundidos na prática empresarial:

- Uma organização **AI-enhanced** usa a IA para melhorar os processos existentes; esse é o modelo predominante atualmente.
- Uma organização **AI-first** a IA projeta seus processos e sua estrutura em torno dos recursos de IA, atribuindo ao julgamento humano apenas as tarefas em que sua vantagem comparativa é inequívoca.
- Uma organização **AI-only** dispensa funcionalmente o trabalho humano em suas operações principais: ela não existe como uma instituição consolidada no momento da elaboração deste relatório, e sua viabilidade em ambientes regulados continua sendo uma hipótese de trabalho.

A situação atual: AI-first como uma fronteira operacional

Os exemplos mais avançados de organizações que priorizam a IA vêm, significativamente, do setor de tecnologia pura, onde a ausência de restrições regulatórias sobre a automação e a natureza digital do produto permitem que o modelo seja levado aos seus limites atuais.

A plataforma de geração de imagens de IA Midjourney gerou mais de US$ 500 milhões em receita em 2025 com uma equipe de aproximadamente 163 pessoas, nenhum investimento em marketing e nenhum financiamento externo[178]. A plataforma de desenvolvimento Cursor (Anysphere) atingiu US$ 500 milhões em receita recorrente anual em maio de 2025, tornando-se a empresa de SaaS de crescimento mais rápido da história, com menos de 50 funcionários[179]. A relação receita por funcionário dessas empresas (mais de US$ 3 milhões em ambos os casos) é uma ordem de grandeza maior do que as referências históricas do setor de tecnologia e dos grandes grupos bancários globais[180].

No campo dos serviços financeiros, um caso avançado é o MYbank, um banco digital chinês de propriedade do Ant Group, que desde 2015 tem operado sob o princípio de zero intervenção humana na aprovação de crédito para PMEs. Seu modelo "310" - três minutos de solicitação, um segundo de aprovação, zero intervenção humana - atendeu a mais de 50 milhões de PMEs. O sistema usa modelos de previsão de fluxo de caixa com mais de 95% de precisão e conta com dados de geolocalização por satélite para avaliação de risco agrícola[181]. O MYbank opera sem uma rede de agências ou força de vendas, embora mantenha equipes de engenharia e gerenciamento: é um modelo que prioriza a AI-First em suas operações principais, não AI-only como um todo.

Por que ainda não existe uma organização AI-only?

O gap entre *AI-first* e *AI-only* não é apenas tecnológica; ela é regulatória, legal e organizacional. Em setores regulados, como o bancário e o de seguros, as normas atuais exigem supervisão e responsabilidade humana pelas decisões relevantes. A Lei Europeia de IA classifica os aplicativos de IA em crédito, saúde e seguro de vida e componentes de infraestrutura crítica como sistemas de

178 Midjourney (2025).
179 Sacra (2025).
180 Dealroom (2025).
181 CKGSB (2025).

alto risco, com requisitos explícitos de supervisão humana[182]. A remoção dessa supervisão no núcleo operacional de uma instituição financeira não é compatível com a estrutura prudencial europeia.

Nos setores não regulados, o limite atual não é regulatório, mas de capacidade. Os agentes autônomos de IA podem executar tarefas complexas, mas sua taxa de erro em fluxos de trabalho estendidos, sua incapacidade de lidar com situações não cobertas pelo treinamento e a ausência de mecanismos de responsabilidade legal equivalentes aos de uma pessoa jurídica significam que a eliminação total do trabalho humano nas operações principais gera riscos operacionais que ainda não são suportáveis[183].

O caso da Klarna ilustra os limites atuais: a empresa reduziu sua força de trabalho de 7.400 para aproximadamente 3.000 entre 2022 e 2025 por meio de um congelamento de contratações e automação extensiva, com um assistente de IA gerenciando o equivalente a 853 funcionários de atendimento ao cliente[184]. Sua trajetória define empiricamente o limite entre onde a IA pode operar de forma autônoma com qualidade suficiente e onde o julgamento humano fornece valor diferencial atualmente.

Previsões dos líderes do setor

Os chefes dos principais laboratórios de IA de ponta fazem previsões que colocam a organização AI-only no horizonte imediato. Sam Altman, CEO da OpenAI, declarou em 2024: "Veremos empresas de dez pessoas com avaliações de bilhões de dólares muito em breve [...] Há uma aposta no meu grupo de bate-papo de amigos executivos sobre quando existirá a primeira empresa de uma pessoa avaliada em um bilhão de dólares, o que seria inimaginável sem a IA. E agora isso vai acontecer"[185]. Quando perguntado em maio de 2025 sobre quando esse cenário se concretizaria, Dario Amodei, CEO da Anthropic, respondeu: "2026"[186].

Amodei desenvolve o argumento em seu artigo de janeiro de 2026, no qual descreve o equivalente funcional de "uma nação de gênios em um data center", ou seja, 50 milhões de agentes mais capazes do que qualquer ganhador do Prêmio Nobel, operando entre dez e cem vezes a velocidade humana; e estima que 50% dos empregos de nível básico poderiam ser cortados dentro de um a cinco anos[187]. O mesmo documento observa que a Anthropic já executa a maior parte do código que produz usando IA, aproximando-se da autonomia operacional total no desenvolvimento de software.

As organizações existentes podem se tornar AI-only?

A questão estratégica subjacente não é se as organizações AI-only existirão, mas como elas passarão a existir. A resposta é contraintuitiva: é improvável que elas surjam a partir da transformação das organizações existentes. Clayton Christensen documentou[188] em The Innovator's Dilemma que as empresas estabelecidas são estruturalmente incapazes de adotar tecnologias disruptivas de dentro para fora: seus processos, incentivos e bases de clientes são otimizados para o modelo estabelecido, e quaisquer iniciativas disruptivas internas competem em desvantagem permanente por recursos e atenção da administração. A transição para a IA em primeiro lugar exacerba essa lógica: uma organização com dezenas de milhares de funcionários tem seus processos projetados para essa escala humana. Esses processos não são redesenhados; eles são substituídos.

O padrão que está surgindo na Ásia aponta para um caminho alternativo: a criação de novas instituições, com sua própria marca e sem patrimônio operacional, que competem livremente até atingirem massa crítica e canibalizarem a empresa original. A Ping An, a maior seguradora do mundo em prêmios emitidos, incubou 11 subsidiárias independentes de tecnologia (incluindo OneConnect, Lufax e Ping An Good Doctor) entre 2013 e 2022, cinco das quais foram listadas como instituições autônomas[189]. O DBS Bank criou o Digibank como um banco digital segregado, operando com um quinto dos recursos por cliente de um banco convencional, e cujo aprendizado alimentou a arquitetura da matriz[190]. O mecanismo é idêntico: uma instituição nova, não legada, que se expande sem as restrições da organização matriz e, se o experimento fracassar, é fechada sem prejudicar a empresa original.

Na Europa e, em menor escala, nos EUA, esse mecanismo encontra atritos estruturais que vão além da regulação da IA. A introdução de sistemas de IA no local de trabalho pode, na Europa, exigir consulta ou negociação com conselhos de trabalhadores e, em alguns países, seu acordo explícito[191]. Os acordos coletivos em setores de emprego intensivo incorporam cláusulas que limitam a automação. A proteção de dados dos funcionários de acordo com o GDPR acrescenta mais complexidade[192]. O resultado é uma assimetria com consequências estratégicas não intencionais: a regulação ocidental dificulta que as organizações existentes criem as instituições que priorizam a IA e que acabariam por desafiá-las.

Quando a tecnologia atingir o limiar da viabilidade de uma organização exclusivamente de IA, a questão de quem a construirá primeiro provavelmente terá uma resposta geográfica.

182 *AI Act* (2024).
183 Amodei (2026).
184 Fortune (2025c).
185 Altman (2024a).
186 Quartz (2025).
187 Amodei (2026).
188 Christensen (1997).
189 IMD (2023).
190 DBS (2024).
191 Baker McKenzie (2025).
192 OIT (2025c).

Gêmeos digitais e simulação do comportamento humano

Da engenharia aeroespacial à simulação universal

O conceito de gêmeo digital tem data e local de nascimento precisos. Em outubro de 2002, Michael Grieves apresentou em um fórum da Society of Manufacturing Engineers o que ele chamou de "Conceptual Ideal for Product Lifecycle Management": a ideia de que qualquer objeto físico poderia ter um correlato digital que o representaria dinamicamente durante todo o seu ciclo de vida, sincronizando em tempo real o estado do objeto real com sua representação virtual. O termo "gêmeo digital" foi posteriormente cunhado por John Vickers, engenheiro-chefe da NASA, que formalizou o conceito no roteiro de tecnologia de 2010 da agência[193]. A definição da NASA nesse documento continua sendo a mais precisa que existe: "uma simulação multifísica, em várias escalas e probabilística de um veículo ou sistema que usa os melhores modelos físicos disponíveis, atualizações de sensores e histórico da frota para espelhar a vida de seu gêmeo físico".

O ponto de partida é importante porque revela a premissa implícita que orienta o desenvolvimento de gêmeos digitais há duas décadas: um gêmeo digital funciona bem quando o sistema que ele modela obedece a leis físicas conhecidas e determinísticas. Uma turbina a gás, a fuselagem de uma aeronave, uma rede elétrica: sistemas complicados, com muitos componentes, mas, em princípio, totalmente modeláveis, desde que haja potência computacional e dados de sensores suficientes. Com base nessa premissa, a tecnologia amadureceu de forma constante. Hoje, os gêmeos digitais de ativos físicos estão em operação, por exemplo, em manufatura avançada, energia, infraestrutura e aviação, com reduções documentadas nos tempos de manutenção não planejada e aceleração substancial dos ciclos de design de produtos.

A fronteira epistemológica: sistemas complicados versus sistemas complexos

A expansão do conceito para além do domínio físico revelou um limite que não é tecnológico, mas epistemológico. Michael Batty, a autoridade acadêmica mais reconhecida em modelagem computacional de cidades, o formula com precisão: os gêmeos digitais trabalham em sistemas complicados (muitas partes, mas comportamento determinável em princípio) e encontram dificuldades estruturais em sistemas complexos, onde o comportamento geral emerge da interação dos agentes e não pode ser deduzido das propriedades de seus componentes individuais[194]. Uma cidade, uma economia, um mercado financeiro, uma organização humana são sistemas complexos nesse sentido técnico preciso.

O filósofo Stefano Moroni desenvolve o argumento em termos ainda mais diretos: as limitações dos gêmeos urbanos digitais não são temporárias (elas não desaparecerão com mais dados ou maior poder computacional), mas derivam da natureza intrinsecamente emergente dos sistemas sociais[195]. A imprevisibilidade dos detalhes em um sistema complexo não é um déficit de informações; é uma propriedade do sistema. Isso tem uma implicação prática imediata: um gêmeo digital de uma fábrica pode prever com alta confiabilidade quando um rolamento irá falhar; um gêmeo digital de uma cidade pode aproximar as tendências agregadas de tráfego, mas não pode prever de forma confiável o efeito de uma política habitacional em padrões de segregação residencial de dez anos. A distinção não é de grau; é de natureza.

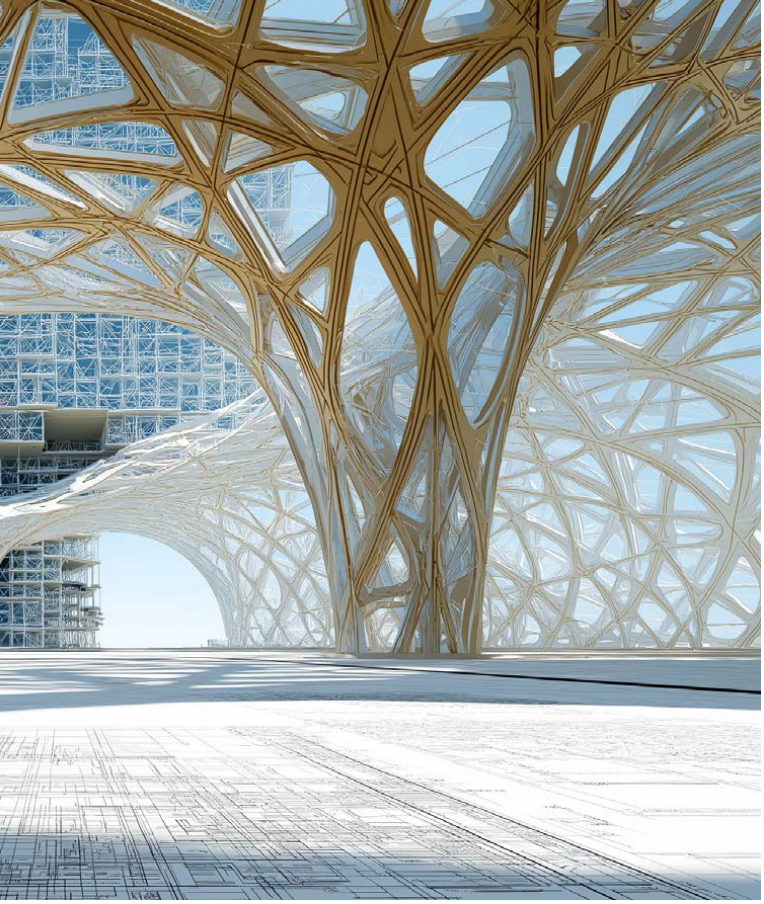
Esse limite epistemológico definiu, durante décadas, o teto do campo de gêmeos digitais. E é exatamente onde os modelos de linguagem em larga escala estão introduzindo uma descontinuidade que merece atenção.

O ponto de inflexão: o comportamento humano se torna modelável

Em abril de 2023, uma equipe de pesquisadores de Stanford publicou um artigo que inaugurou uma linha de trabalho radicalmente nova. Joon Sung Park e seus coautores criaram 25 agentes computacionais (cada um deles dotado de uma identidade, uma memória persistente, um conjunto de relações sociais e uma capacidade de raciocínio baseada em um modelo de linguagem) e os colocaram em um ambiente simulado equivalente a uma pequena cidade[196]. Os agentes acordaram, tomaram café da manhã, foram trabalhar, formaram opiniões, iniciaram conversas e coordenaram atividades coletivas sem que esses comportamentos fossem explicitamente programados: eles surgiram da interação entre a memória individual de cada agente, sua capacidade de refletir sobre experiências passadas e seu modelo do ambiente social. O artigo ganhou o Best Paper Award no ACM Symposium on User Interface Software and Technology em 2023. A comunidade científica reconheceu que algo qualitativamente novo havia acontecido.

193 Grieves-Vickers (2017).
194 Batty (2024).
195 Moroni (2025).

196 Park (2023).



O motivo subjacente é que os modelos de linguagem em larga escala absorveram, durante o treinamento, uma quantidade extraordinária de comportamento humano registrado: conversas, decisões, padrões de raciocínio, respostas emocionais, normas sociais implícitas. Eles não aprenderam explicitamente as leis do comportamento humano (ninguém as conhece com essa precisão), mas desenvolveram uma aproximação estatisticamente densa que, sob condições controladas, gera um comportamento plausível. Pela primeira vez, a premissa que impedia a modelagem de sistemas sociais complexos foi parcialmente eliminada: não porque o comportamento humano tenha deixado de ser emergente, mas porque agora existe um gerador de comportamento suficientemente rico para preencher uma simulação com agentes confiáveis.

A extensão natural desse trabalho surgiu em novembro de 2024. A mesma equipe de Stanford publicou os resultados de um experimento de escala diferente: 1.052 pessoas reais, entrevistadas em profundidade sobre suas vidas, atitudes e experiências, foram transformadas em agentes que replicam suas respostas e comportamentos em pesquisas padronizadas e experimentos sociais. Os agentes geradores replicaram as respostas de indivíduos reais na General Social Survey com 85% de precisão (estatisticamente comparável à variabilidade natural do próprio indivíduo ao responder à mesma pesquisa duas semanas depois) e obtiveram resultados comparáveis na previsão de traços de personalidade e em experimentos de ciências sociais[197]. O que em 2023 era uma demonstração conceitual com personagens fictícios, em 2024 tornou-se uma metodologia empiricamente validada com pessoas reais.

197 Park (2024).

Aplicações atuais e horizonte futuro

As implicações desse salto são intersetoriais. Na pesquisa de mercado, a startup Simile - fundada por Joon Sung Park juntamente com Michael Bernstein e Percy Liang, coautores do paper fundacional, e apoiada em fevereiro de 2026 com US$ 100 milhões pela Index Ventures com a participação de Fei-Fei Li e Andrej Karpathy - cria gêmeos digitais de pessoas reais para ajudar as empresas a simular o comportamento do cliente antes de lançar um produto, modificar uma política de preços ou redesenhar uma experiência do usuário[198]. Em uma demonstração pública, a plataforma previu corretamente oito das dez perguntas feitas por analistas em uma chamada simulada[199]. O setor de pesquisa, de mercado global de US$ 142 bilhões[200], está enfrentando uma ruptura estrutural: o que hoje exige semanas de trabalho de campo pode ser executado em horas em populações sintéticas.

Na política pública e no planejamento urbano, um uso mais sutil, mas igualmente transformador, é articulado: não o gêmeo digital como um oráculo preditivo, mas como um laboratório de cenários em que as consequências de diferentes intervenções podem ser exploradas antes de se comprometer recursos reais[201]. Na regulação financeira, essa abordagem tem aplicação direta no teste de estresse de cenários macroeconômicos adversos e na simulação do comportamento do mercado em face de intervenções regulatórias.

A trajetória do campo aponta para uma escala qualitativamente diferente. Se hoje é possível simular com alta fidelidade mil pessoas reais, a questão no horizonte imediato é o que acontecerá quando esse número chegar a um milhão, cem milhões, uma sociedade inteira modelada em tempo real. As aplicações em políticas públicas, regulação econômica e projeto institucional serão de uma ordem de grandeza diferente da pesquisa de mercado: não antecipar qual produto um consumidor comprará, mas prever como uma população responderá a uma reforma tributária, uma crise de saúde ou uma mudança na política monetária antes que essa intervenção seja implementada no mundo real. Essa capacidade não tem precedentes históricos nem, até o momento, nenhuma estrutura de governança para regulamentá-la.

198 Index Ventures (2026).
199 Bloomberg (2026).
200 ESOMAR (2024).
201 Bettencourt (2024).

Ambient AI e computação invisível

A interface é o entorno

Ambient AI - ou inteligência de ambiente - é a IA que opera sem ser invocada. Diferentemente dos sistemas convencionais, que respondem a uma instrução explícita do usuário, os sistemas de ambiente observam continuamente o contexto, inferem necessidades e agem proativamente. A interface desaparece não porque tenha sido aprimorada, mas porque o sistema não precisa mais dela: o próprio ambiente se torna o ponto de interação. A computação torna-se "invisível" no sentido literal: incorporada a objetos, espaços e processos sem que o usuário a perceba como tal[202].

Essa inversão (do usuário indo até o sistema para o sistema vindo até o usuário) é possível hoje pela convergência de três desenvolvimentos simultâneos: a miniaturização de modelos capazes de serem executados em dispositivos edge, mantendo um estado contínuo – memória do usuário cumulativa atualizada localmente – sem depender de conectividade em nuvem (edge AI e TinyML), a densificação de redes de sensores físicos e biométricos e a capacidade dos LLMs de raciocinar sobre contextos heterogêneos e ambíguos em tempo real[203]. Nenhum dos três é novo por si só; é a maturidade simultânea deles que permite que a Ambient AI passe do conceito à implantação operacional.

Estado atual da implementação

O exemplo mais documentado de Ambient AI em operação são os ambient AI scribes em ambientes clínicos: sistemas que ouvem continuamente a conversa entre o paciente e o médico, inferem o contexto clínico sem instruções explícitas e geram automaticamente a documentação do encontro. O ensaio clínico randomizado da UCLA avaliou duas plataformas (Microsoft DAX e Nabla) em 238 médicos de 14 especialidades e mais de 72.000 encontros: reduziu-se a carga de documentação e melhorou os indicadores de esgotamento profissional[204].

O sistema não foi invocado iterativamente durante a consulta: ele ouviu, inferiu e escreveu. Ainda é uma forma restrita de inteligência ambiental (contexto delimitado, propósito claro, episódio definido). A Ambient AI madura não operará dentro da consulta, mas na escala de todo o hospital, correlacionando padrões longitudinais sem nenhum ponto inicial ou final determinado pelo usuário.

O futuro físico e digital

As implementações de hoje são a ponta de uma transformação mais ampla. Nos próximos anos, a Ambient AI se estenderá a ambientes físicos e digitais e fará com que os casos atuais pareçam rudimentares:

Ambientes físicos

- **Espaços de trabalho adaptáveis**. O ambiente infere o estado de atenção do ocupante a partir da frequência cardíaca, da variabilidade do ritmo, dos padrões de movimento e reconfigura a temperatura, a luz e o nível de ruído para otimizar o desempenho cognitivo sem a intervenção consciente do usuário.
- **Manutenção industrial antecipatória**. Os sistemas não alertarão quando o equipamento falhar: eles detectarão o padrão de comportamento que precede a falha com antecedência suficiente para reorganizar a produção. O evento disruptivo desaparece do horizonte operacional.

202 Bimpas (2024); Hernandez-Torres (2025).
203 Heydari (2025).
204 Lukac (2025).

- **Wearables com perfil individual**. Os dispositivos da próxima geração compararão os sinais vitais do usuário não com as médias da população, mas com seu próprio histórico fisiológico. O alerta será acionado antes que o usuário esteja ciente do sintoma.
- **Infraestrutura urbana reativa**. Redes de transporte, iluminação e gestão de resíduos que se autoajustam em tempo real aos padrões de uso inferidos, sem planejamento centralizado explícito ou intervenção humana no circuito.
- **Assistência domiciliar invisível**. Sistemas que monitoram continuamente pessoas idosas ou com doenças crônicas, detectam anomalias nas rotinas (padrões de sono, mobilidade, alimentação) e ativam protocolos de alerta ou intervenção sem a solicitação do usuário.

Ambientes digitais

- **Ambientes de desenvolvimento que antecipam o problema**. Os assistentes de programação proativos evoluirão para sistemas que, antes que o desenvolvedor identifique o erro, terão mapeado o espaço de soluções prováveis e apresentado opções no momento cognitivamente apropriado[205].
- **Gestão da atenção, não apenas da informação**. Os sistemas não fornecerão informações quando elas estiverem disponíveis, mas quando o usuário estiver em condição de processá-las: modelando o estado de atenção ao longo do dia e avaliando o momento da interrupção.
- **Contexto organizacional contínuo**. Sistemas que conhecem a todo momento o status dos projetos, as comunicações pendentes e as decisões em andamento, além de apresentar proativamente informações relevantes a cada membro da equipe sem que sejam solicitadas.
- **Negociação autônoma de recursos**. Agentes ambientais que gerenciam em nome do usuário (cronograma, orçamento, acesso a serviços) dentro de parâmetros definidos, sem exigir aprovação explícita para cada decisão de baixa complexidade.

Implicações que a tecnologia não resolve

A Ambient AI não levanta apenas questões de privacidade. Ela gera um conjunto mais amplo de tensões que as estruturas de governança atuais não resolveram.

A primeira é a natureza do erro. Em um sistema invocado, o erro é visível: o usuário solicitou algo, o sistema respondeu mal. Em um sistema ambiente, o erro pode ser um erro de segurança. Em um sistema ambiente, o erro pode não ser percebido porque não houve solicitação explícita com a qual comparar a resposta. O ambient scribe da UCLA registrou imprecisões clinicamente significativas em uma proporção de encontros[206:] em um sistema invisível, o mecanismo de detecção de erros precisa ser deliberadamente projetado, pois não surge naturalmente da interação.

A segunda é a assimetria de poder entre aquele que projeta o ambiente e aquele que o habita. Em um hospital, um escritório ou um prédio público, o usuário não escolhe se o ambiente é inteligente: ele habita um espaço cujas inferências sobre seu comportamento foram moldadas por terceiros. Tshilidzi Marwala, chanceler da Universidade das Nações Unidas, coloca com precisão: A Ambient AI tem um apetite por dados – íntimos, comportamentais, biométricos – que torna as noções convencionais de consentimento informado estruturalmente inadequadas[207]. O *AI Act Europeu*, projetado para sistemas invocados com funções delimitadas, não aborda de forma satisfatória esses ambientes de observação contínua.

A terceira é a dependência cognitiva. Um sistema que gerencia proativamente a atenção do usuário, o fluxo de informações e as interrupções não apenas auxilia o trabalho do usuário, mas também molda a arquitetura cognitiva do usuário. A pergunta feita em 2003, "a computação com reconhecimento de contexto está tirando o controle do usuário?"[208] ficou sem resposta por décadas. A escala em que a Ambient AI a coloca hoje transforma o que era uma questão acadêmica em um problema de design com consequências operacionais imediatas.

A quarta tensão é a responsabilidade causal. Nos sistemas invocados, a rastreabilidade é relativamente simples: há uma instrução, uma resposta, um momento de decisão atribuível. Nos sistemas de ambiente, a cadeia causal é confusa. Se um sistema de manutenção antecipada reorganiza a produção e essa reorganização condiciona as decisões humanas subsequentes, o limite entre a atuação técnica e a atuação humana não é claro. A regulação atual, incluindo o *AI Act*, pressupõe a finalidade pretendida e a avaliação de risco ex ante; a IA ambiental introduz a finalidade emergente e o comportamento adaptativo contínuo, o que sobrecarrega diretamente os mecanismos de compliance existentes.

A Ambient AI não muda apenas a forma como trabalhamos ou como cuidamos de nós mesmos: ela muda a sequência entre necessidade e consciência. Um sistema pode saber do que precisamos antes de sabermos. Se esse recurso for implantado na escala que a trajetória do campo sugere, as questões que ele levanta poderão ir além da tecnologia e da regulação, chegando a algo mais fundamental: o que significa tomar as próprias decisões em um ambiente que já as antecipou.

205 Chen (2025); Pu (2025).
206 Lukac (2025).
207 Marwala (2025).
208 Barkhuus (2003).

Interação entre IA e computação quântica

Duas tecnologias diferentes, uma interseção que importa

A IA e a computação quântica são tecnologias distintas com princípios, horizontes de tempo e casos de uso completamente diferentes. A IA já está operacional em escala industrial; a computação quântica ainda é, em sua maior parte, um campo de pesquisa avançado com implementações muito limitadas. Sua interação, em ambas as direções, tem implicações concretas para qualquer organização que dependa de sistemas digitais.

Um computador clássico resolve problemas testando opções sequencialmente ou em paralelo, mas sempre dentro de um espaço de possibilidades que cresce de forma gerenciável. Há problemas para os quais isso não é suficiente: otimizações com milhares de variáveis interdependentes, simulações de sistemas moleculares ou determinados problemas matemáticos cuja dificuldade é justamente a base da criptografia moderna. Um computador quântico opera de uma maneira radicalmente diferente: em vez de testar as opções uma a uma, ele pode explorar simultaneamente um espaço de possibilidades de uma dimensionalidade que nenhum sistema clássico pode representar. Para essa classe específica de problemas - não todos - a diferença de desempenho não é incremental, mas de uma ordem de magnitude maior.

O problema é que a construção de um computador quântico que funcione de forma confiável tem se mostrado extraordinariamente difícil. As informações quânticas são extremamente sensíveis a perturbações ambientais - temperatura, vibrações, interferência eletromagnética - e os erros se acumulam rapidamente. Durante décadas, o campo avançou muito mais rapidamente na teoria do que no hardware. Isso mudou parcialmente em dezembro de 2024.

O marco de 2024 e o que ele significa

O Google publicou os resultados de seu processador Willow na Nature: o primeiro sistema a mostrar que, à medida que mais componentes computacionais são adicionados, os erros diminuem em vez de aumentar[209]. Esse é um resultado que a teoria previa desde 1995, mas que nenhum sistema havia conseguido realizar. A importância não está nos números de desempenho (que são impressionantes, mas em benchmarks artificiais), mas no que isso implica para a trajetória do campo: o obstáculo que por trinta anos impediu que esses sistemas fossem escalonados de forma confiável foi superado no laboratório.

A distância entre essa conquista e um computador quântico com aplicações comerciais ainda é considerável. Os próprios pesquisadores do Google colocam esse horizonte em torno do final da década. Mas a direção não está mais em discussão: o problema central estava na correção de erros, e esse problema agora tem uma solução comprovada. O que resta é a engenharia para escala, não um salto científico no vazio.

209 Google Quantum AI (2024).

Três maneiras pelas quais isso afeta a IA

A interseção entre a computação quântica e a IA opera em três planos distintos, com urgências distintas.

O primeiro é a aceleração do machine learning. O treinamento de um modelo de IA em larga escala é, em sua essência, um problema de otimização matemática em espaços de dimensões enormes: encontrar os valores de bilhões de parâmetros que minimizem o erro de previsão. Esse é exatamente o tipo de problema para o qual a computação quântica oferece uma vantagem teórica. Foi demonstrado formalmente[210] que os sistemas quânticos tolerantes a falhas poderiam acelerar substancialmente os algoritmos de treinamento de modelos em grande escala, reduzindo o tempo computacional e o consumo de energia. Para desenvolver isso, é necessário hardware que ainda não existe em escala suficiente. Mas quando estiver disponível, poderá alterar radicalmente a economia do treinamento de modelos de IA, hoje dominada por aqueles que podem pagar por uma infraestrutura de GPU em grande escala.

O segundo plano é o próprio *machine learning* quântico**:** usar processadores quânticos para executar algoritmos de *machine learning* com mais eficiência. Aqui, a literatura é mais cautelosa, descrevendo[211] tanto as promessas quanto os obstáculos reais: a vantagem quântica nas tarefas de aprendizado não é universal nem automática e, em muitos casos, os sistemas clássicos com acesso a dados são competitivos em relação aos quânticos, mesmo em problemas projetados para favorecer os últimos. Em outras palavras: os dados, bem utilizados, podem compensar a vantagem quântica em muitos regimes[212]. O hype sobre a IA quântica como um acelerador universal está à frente das evidências; as aplicações reais serão específicas, não transversais.

O terceiro plano inverte a direção da influência: não é a computação quântica a serviço da IA, mas a computação quântica como uma ameaça à infraestrutura de segurança na qual todos os sistemas digitais, inclusive os sistemas de IA, operam. Toda a criptografia que protege as comunicações digitais atualmente (transações bancárias, autenticação de identidade, canais seguros entre sistemas) baseia-se em problemas matemáticos cuja dificuldade é considerada inatacável para os computadores clássicos. Um computador quântico suficientemente poderoso os resolveria diretamente. Esse nível é o mais urgente, porque parte de seus efeitos já está presente.

A ameaça que já está ativa

A estratégia conhecida como "harvest now, decrypt later" consiste em capturar comunicações criptografadas hoje com a intenção de descriptografá-las quando a computação quântica amadurecer o suficiente. Atores estatais com recursos avançados de inteligência vêm aplicando essa estratégia há anos[213]. Os dados que exigem décadas de confidencialidade (registros médicos, segredos comerciais, comunicações regulatórias, informações financeiras confidenciais) estão sendo comprometidos agora, independentemente de quando chegará o sistema quântico capaz de descriptografá-los.

210 Liu (2024).
211 Li (2025).
212 Huang (2021).
213 Mascelli (2025).

A resposta institucional mais rigorosa é a do NIST dos EUA, que em agosto de 2024 – após oito anos de trabalho e mais de oitenta propostas de equipes de pesquisa de todo o mundo – publicou os três primeiros padrões de criptografia resistentes a ataques quânticos[214]. Os novos algoritmos são baseados em estruturas matemáticas para as quais não se conhece nenhum ataque quântico eficiente. O NIST recomenda que as organizações iniciem a migração imediatamente; os sistemas federais dos EUA têm um prazo até 2035. Para as instituições financeiras com dados de longa duração, esse prazo não é o horizonte para o início da transição: é o limite para sua conclusão.

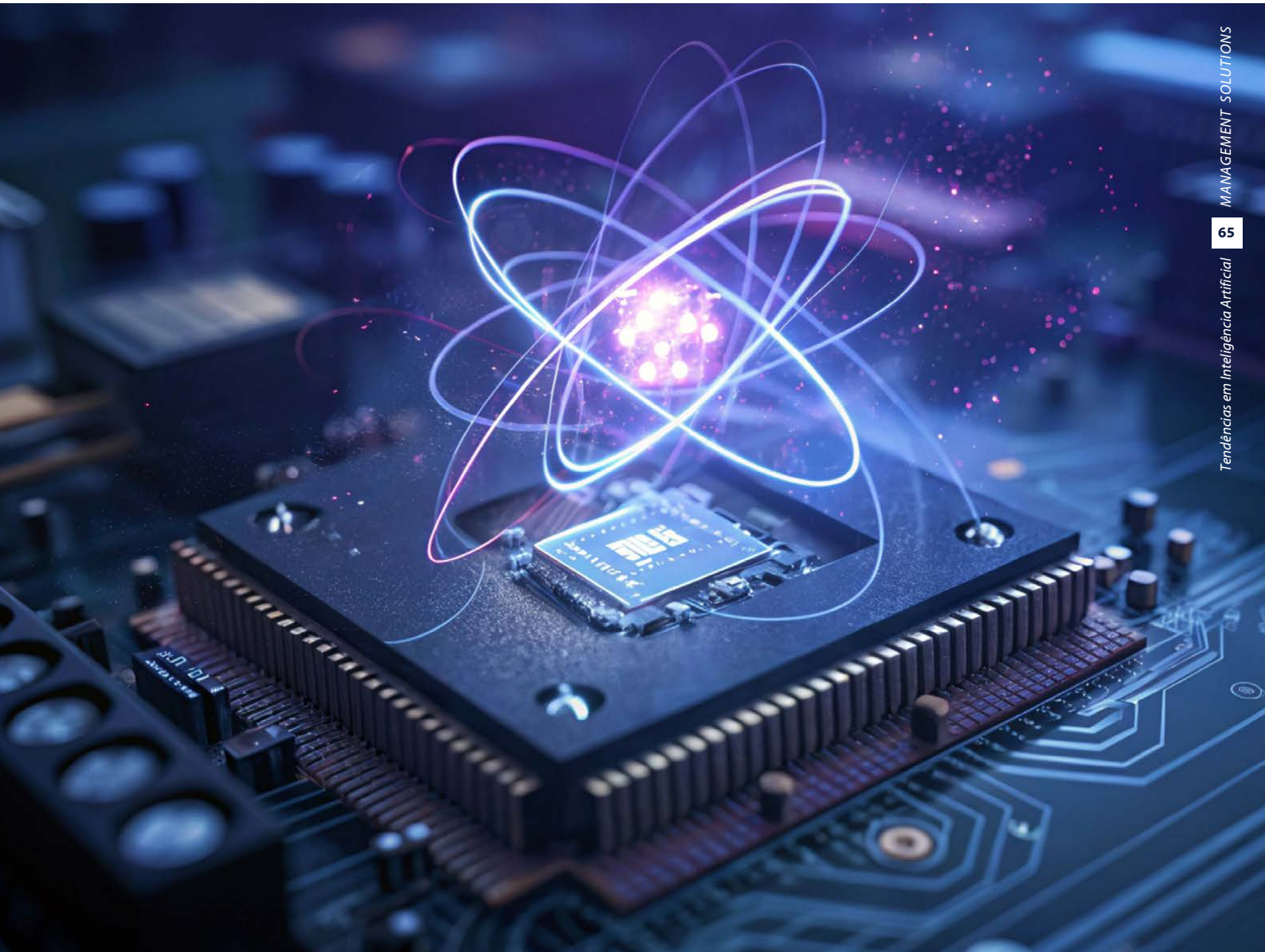
O horizonte da próxima década

Nos próximos anos, a interseção entre IA e computação quântica deixará de ser um tópico de previsão tecnológica para se tornar um elemento com impacto operacional em duas frentes simultâneas.

A frente ofensiva - a aceleração dos recursos de IA – virá com a maturidade do hardware tolerante a falhas, prevista para o final da década. O acesso inicial será feito por meio de serviços em nuvem, replicando a trajetória que a computação de alto desempenho seguiu com as GPUs: primeiro acessível apenas aos mais bem financiados, depois democratizada pela concorrência entre os fornecedores. As organizações que desenvolveram competências de IA até então estarão mais bem posicionadas para aproveitar essa aceleração quando ela chegar.

A frente defensiva (migração de criptografia) não pode ser adiada. A janela de prontidão se estreita à medida que os processadores quânticos aumentam de escala, e a migração de infraestruturas criptográficas em organizações complexas leva anos. O inventário de ativos vulneráveis, a priorização por tempo de vida dos dados e o planejamento da transição são tarefas que devem começar agora, e não quando o computador quântico relevante estiver operacional e for tarde demais.

214 NIST (2024b).



Inteligência Artificial Geral (AGI) como um horizonte estratégico

O que é AGI, ela já chegou?

O termo "inteligência geral artificial" (AGI) apareceu pela primeira vez em 1997, em uma conferência sobre nanotecnologia em Palo Alto. Seu autor, Mark Gubrud, candidato a PhD na Universidade de Maryland, não estava descrevendo uma meta tecnológica desejável: ele estava alertando sobre um risco. Em seu artigo "Nanotechnology and International Security" (Nanotecnologia e segurança internacional)[215], Gubrud definiu AGI como sistemas capazes de "rivalizar ou superar o cérebro humano em complexidade e velocidade, adquirir, manipular e raciocinar com conhecimento geral e ser utilizável em qualquer fase de operações em que a inteligência humana seria necessária". A definição passou despercebida por quase uma década, até que Ben Goertzel e Shane Legg a resgataram e a popularizaram como um rótulo técnico no título de seu livro coletivo Artificial General Intelligence[216]. O termo nasceu como um conto de advertência, mas se tornou uma missão.

Em fevereiro de 2026, a Nature publicou quase simultaneamente dois textos que cristalizaram o debate mais relevante da tecnologia contemporânea. O primeiro, assinado por quatro acadêmicos da UC San Diego[217] - filosofia, *machine learning*, linguística e ciência cognitiva – afirma inequivocamente que a AGI já existe: os LLMs de hoje passam no teste de Turing, ganham medalhas de ouro em olimpíadas de matemática e colaboram na comprovação de teoremas. O segundo[218], publicado quinze dias depois como uma correspondência na mesma revista, responde que essa conclusão só é possível redefinindo o conceito de forma irreconhecível: a definição clássica de AGI formulada em 2007 exige robustez sob novidade, generalização transferível e autonomia de metas, e os sistemas atuais não atendem a esses requisitos. O fato de os principais pesquisadores, com acesso aos mesmos sistemas e dados, chegarem a conclusões opostas reflete o fato de que "inteligência geral" é um conceito contínuo sem limites precisos.

A própria Fei-Fei Li, que construiu os fundamentos da visão computacional moderna e trabalha com inteligência espacial, admite prontamente: "Para ser sincera, tenho dificuldades com essa definição de AGI"[219]. Dario Amodei, CEO da Anthropic, vai além: ele declara abertamente que não gosta do termo e prefere falar sobre "IA poderosa": sistemas com capacidades intelectuais comparáveis ou superiores às de um ganhador do Prêmio Nobel na maioria das disciplinas[220].

Essa é a perspectiva correta para as organizações. A questão estrategicamente relevante não é filosófica - será que chegamos à AGI – mas funcional: quando um sistema poderá executar de forma totalmente autônoma ciclos completos de trabalho de alto valor cognitivo em todos os domínios? Esse limite já foi ultrapassado em vários setores.

O estado atual: brilhantismo e fragilidade simultâneos

Andrej Karpathy, cofundador da OpenAI e ex-diretor de IA da Tesla, cunhou o conceito de "inteligência irregular" para descrever a condição atual dos LLMs: sistemas que resolvem problemas de olimpíadas matemáticas e não conseguem determinar qual número é maior, 9,11 ou 9,9; que são fluentes em dezenas de idiomas e sofrem do que ele chama de "amnésia anterógrada"[221], incapacidade de consolidar o aprendizado entre as sessões. Fei-Fei Li os descreve[222] como “wordsmiths in the dark: eloquent but inexperienced, knowledgeable but ungrounded”: eloquentes, mas sem experiência incorporada no mundo físico.

E, no entanto, esses mesmos sistemas já redigem contratos, analisam o risco de crédito, sintetizam a literatura científica, geram e auditam códigos complexos, produzem resumos regulatórios. Eles fazem isso de forma autônoma, em uma velocidade e escala que nenhum computador humano pode igualar. A questão não é mais quando essa capacidade chegará. É o que faremos com o que já chegou e como nos prepararemos para o que virá em seguida.

A cadeia causal: o que virá depois

A transição em andamento segue uma lógica de escalonamento cumulativo. O primeiro movimento é da ferramenta para o agente: os sistemas passam da resposta às instruções para a busca de metas por meio de ciclos autônomos de ação, observação e correção. O segundo movimento é do agente para a infraestrutura ambiental. Karpathy o formula com precisão em[223]: os LLMs são a nova eletricidade. A IA deixa de ser uma tecnologia que "adotamos" e passa a ser uma tecnologia que "acontece conosco": tecido operacional invisível dos sistemas que usamos, uma condição do ambiente em vez de uma ferramenta nele.

O terceiro movimento é o mais subestimado pela análise convencional: o loop recursivo. Os modelos já são usados para aprimorar modelos, gerando dados de treinamento sintéticos, otimizando arquiteturas e produzindo hipóteses de pesquisa. Isso cria uma dinâmica em que a velocidade do aprimoramento da IA é uma função da inteligência da IA. Amodei chama isso de "o fim do exponencial": não o ponto em que a curva se achata, mas onde o vetor de aceleração se torna superexponencial, porque o agente de aceleração é o sistema que está sendo acelerado[224].

A consequência estrutural desse loop não tem precedentes na história da civilização: o limite superior de raciocínio disponível no planeta tem sido, desde os primeiros hominídeos, a inteligência humana. No momento, esse limite está sendo deslocado em domínios específicos. Quando a mudança se tornar geral e robusta, a taxa de mudança tecnológica e científica se tornará parcialmente dissociada da capacidade humana de compreensão e verificação. Essa não é uma projeção apocalíptica, mas uma descrição estrutural do que esse ciclo recursivo implica.

215 Gubrud (1997).
216 Goertzel, Legg (2007).
217 Chen, Belkin, Bergen, Danks (2026).
218 Quattrociocchi, Capraro, Marcus (2026).
219 Li (2025).
220 Amodei (2024b).
221 Karpathy (2025).
222 Li (2025).
223 Karpathy (2025).
224 Amodei (2024a).

O gap de absorção

Há duas curvas que avançam em velocidades radicalmente diferentes. A curva técnica – exponencial, que se acelera automaticamente por meio do loop recursivo - comprime em anos o que costumava levar décadas. A curva de absorção organizacional – redesenho de processos, reengenharia de funções, construção de infraestrutura de governança, gerenciamento de mudanças institucionais – avança mais lentamente, com atritos consideráveis: sistemas legados, dificuldades organizacionais, resistência cultural, regulação que chega tarde, escassez de talentos capazes de integrar esses recursos em operações reais.

O gap entre as duas curvas é a variável definidora da próxima década. A vantagem competitiva não virá do acesso aos melhores modelos, que se tornarão progressivamente comoditizados, nem de seu custo, que continuará a cair exponencialmente. Ela virá da velocidade e do rigor com que uma organização é capaz de se redesenhar para operar com agentes autônomos de forma eficaz e responsável. Esse padrão está documentado empiricamente[225]: os ganhos de produtividade mais significativos não aparecem quando a IA substitui tarefas, mas quando ela reorganiza processos inteiros e redefine a colaboração homem-máquina.

A redefinição do papel da pessoa

A pergunta certa, portanto, não é quais empregos desaparecerão, mas o que pesssoa faz que um sistema de IA não pode fazer, mesmo que seja mais rápido, mais barato e mais consistente. A resposta usual (criatividade, empatia, liderança) é verdadeira, mas insuficiente. Há dimensões que a análise convencional subestima:

- Responsabilidade com consequências reais. Os sistemas de IA não podem ser levados ao tribunal ou perder uma reputação construída ao longo de décadas. Em ambientes financeiros, de saúde, jurídicos e regulatórios, a presença humana é um requisito estrutural.
- Certificação e fé pública. O tabelião que certifica, o médico que assina, o auditor que contra-assina: esses atos valem não por causa do processamento de informações envolvido, mas por causa da responsabilidade pessoal e institucional por trás deles.
- Relacionamento interpessoal. Há contextos em que o que é necessário não é a resposta certa, mas a presença de outra pessoa: dor, conflito, cuidado. Substituir essa presença por um sistema mais eficiente não resolve o problema.
- Formulação de perguntas que valem a pena e juízo moral. Quando a IA faz bem o que lhe é pedido, o valor muda para quem decide o que perguntar: quem define as metas, quem reconhece quais problemas merecem atenção, quem decide quando há valores conflitantes.
- Legitimidade democrática. As decisões que afetam as comunidades exigem deliberação e responsabilidade que não podem ser delegadas a sistemas opacos, por mais precisos que sejam.

O que essas dimensões têm em comum aponta para algo mais profundo do que uma distribuição de tarefas. Ao longo da história, o ser humano tem sido simultaneamente o agente que produz e o sujeito que é responsável pelo que é produzido: é essa unidade entre ação e responsabilidade que os sistemas de IA dissociam estruturalmente. O que é redefinido, então, não é apenas o que fazemos, mas onde residimos na cadeia causal: cada vez menos sobre execução, cada vez mais sobre intenção, julgamento e responsabilidade.

O verdadeiro risco sistêmico: concentração

O risco sistêmico emergente não é o da máquina em revolta. É o da concentração sem precedentes da capacidade cognitiva em um pequeno número de atores – laboratórios, corporações, estados – cuja vantagem se autoamplifica por meio do mesmo ciclo recursivo que acelera o progresso geral. Amodei escreve que, se um estado autoritário conseguisse obter domínio ofensivo em segurança cibernética ou biologia à frente de todos os outros graças à IA, as consequências geopolíticas seriam assimétricas e irreversíveis. Hassabis propõe, em resposta, um modelo de colaboração internacional inspirado no CERN: governança multilateral das etapas finais em direção a sistemas gerais de IA[226].

Atualmente, a resposta institucional (regulatória, corporativa, internacional) está muito aquém da velocidade do problema. A AGI como um horizonte estratégico não exige que as organizações resolvam o debate filosófico sobre se ela já chegou. Exige que elas ajam com a consciência de que suas consequências já estão ocorrendo: nos sistemas que operam hoje, nos ciclos de trabalho que estão sendo redefinidos hoje, nas decisões de governança que estão sendo tomadas ou contornadas hoje.

225 Mollick (2024).

226 Amodei, Hassabis (2026).

07 | Estudo de caso: GenMS™ Sybil

«Falar é grátis. Mostre-me o código».

Linus Torvalds

O GenMS™ Sybil foi especificado, criado, protegido, validado e implantado em um único dia de trabalho[228].

Este caso documenta como.

227 Linus Torvalds (n. 1969), engenheiro de software finlandês-americano, criador do kernel Linux e do Git, dois dos projetos de software livre mais influentes da história.

228 O escopo é deliberadamente restrito: um assistente conversacional de corpus fechado, sem integrações com sistemas corporativos, sem persistência de sessão e com uma superfície de ataque limitada. Essa restrição não é uma simplificação; é uma decisão de projeto. A afirmação não sugere que sistemas de maior complexidade exijam o mesmo esforço: a complexidade e o tempo de desenvolvimento podem aumentar de forma não linear com o número de integrações, usuários simultâneos, requisitos regulatórios e criticidade operacional.

O sistema

O GenMS™ Sybil é um assistente de conversação de acesso público, desenvolvido com base no conteúdo completo deste documento. Ele responde a perguntas, explora implicações e acompanha a reflexão sobre as tendências discutidas aqui. Não armazena dados pessoais dos usuários e, portanto, não apresenta dificuldades no que diz respeito ao de acordo com o Regulamento Geral de Proteção de Dados. O sistema está em compliance desde o desenho: a classificação regulatória, os requisitos do *AI Act* e as obrigações de privacidade não foram incorporados como uma camada subsequente de compliance, mas como critérios de projeto desde a primeira fase de especificação.

O processo: ciclo de vida do LLMOps

A construção seguiu as fases do ciclo de vida do LLMOps sequencialmente e sem exceções.

Preparação dos dados. O corpus do GenMS™ Sybil é este documento. A decisão de não estender o sistema com o conteúdo completo das fontes citadas foi deliberada: fazer isso teria introduzido riscos de direitos autorais e de propriedade intelectual que seriam difíceis de gerenciar. O GenMS™ Sybil está ciente das referências, citações e links para elas, mas não reproduz seu conteúdo. O controle e a minimização de fontes são aqui tanto uma decisão técnica quanto um requisito de compliance.

Experimentação e desenvolvimento. Essa fase produziu a especificação completa do sistema: arquitetura, comportamento esperado, taxonomia de casos de uso, limites operacionais, critérios de qualidade e requisitos de segurança. A especificação, com dezenas de páginas, foi construída em diálogo com um LLM por meio de *vibe coding*: o profissional formulou objetivos, avaliou propostas e tomou decisões; a máquina materializou a intenção na documentação técnica de produção. Configurações alternativas de modelos foram avaliadas, versões prontas foram gerenciadas desde o início e métricas de avaliação qualitativa foram definidas: coerência, factualidade, adequação contextual, comportamento diante de perguntas fora do escopo.

Validação. A avaliação do GenMS™ Sybil integrou a revisão humana ao processo, o teste de estresse semântico e os exercícios de red-teaming com o objetivo de identificar comportamentos indesejados. A validação não foi um evento isolado no final do processo; ela foi contínua durante todo o ciclo. O GenMS™ Atlas - sistema da Management Solutions para testar sistemas baseados em LLM - avaliou o sistema em várias de suas 26 dimensões: viés, consistência, privacidade, robustez, explicabilidade e compliance regulatório. Os problemas detectados foram resolvidos antes da implementação; aqueles que persistem são documentados.

Implementação. A construção do sistema foi executada pela Claude Code com base na especificação completa. O resultado foi um aplicativo coerente, com lógica de contexto, gerenciamento de sessão e interface de usuário. O código foi auditado em busca de vulnerabilidades e vetores de ataque; as correções foram incorporadas no mesmo ciclo. A implantação levou em conta as implicações de infraestrutura, latência e custo de um sistema generativo em produção desde o início.

Monitoramento. O GenMS™ Sybil opera com monitoramento ativo dos custos por token, rastreabilidade total das interações para auditoria e supervisão regulatória, além de alertas para comportamentos anômalos ou padrões de uso imprevistos. O processo de desenvolvimento foi iterativo: a primeira versão não era a versão final. A iteração controlada, com critérios de avaliação explícitos em cada ciclo, é o que distingue a industrialização da experimentação.

Decisões de arquitetura

O projeto do GenMS™ Sybil envolveu dilemas técnicos concretos, resolvidos com critérios explícitos:

- RAG vs. contexto completo: contexto completo. O modelo subjacente tem uma janela de entrada de um milhão de tokens; o documento cabe em sua totalidade em cada conversa. A fragmentação do RAG destruiria a coerência geral que as perguntas mais valiosas exigem, e o argumento de custo que historicamente o justificou não compensa mais essa deterioração da qualidade.
- Extensão do corpus com fontes citadas: descartada. O risco de violação de direitos autorais e de propriedade intelectual é incompatível com um sistema de acesso público. O GenMS™ Sybil cita e vincula referências; ele não reproduz seu conteúdo.
- Ajuste fino vs. solicitação: solicitação com documento no contexto. O ajuste fino é caro, lento e opaco às atualizações de documentos. O *prompt*ing garante a rastreabilidade total de cada mudança de comportamento.
- Modelo proprietário vs. modelo de código aberto: modelo de fronteira proprietário. A maturidade dos modelos de código aberto é insuficiente para garantir a consistência e os controles de segurança necessários em um sistema público sem supervisão humana contínua.
- Memória persistente vs. sessões independentes: sessões independentes. Minimização de dados, compliance estrutural com o GDPR e eliminação do risco de contaminação entre conversas.
- Acesso aberto vs. registro: registro mínimo com nome, organização e e-mail. Controle de acesso viável e estatísticas de uso agregadas.

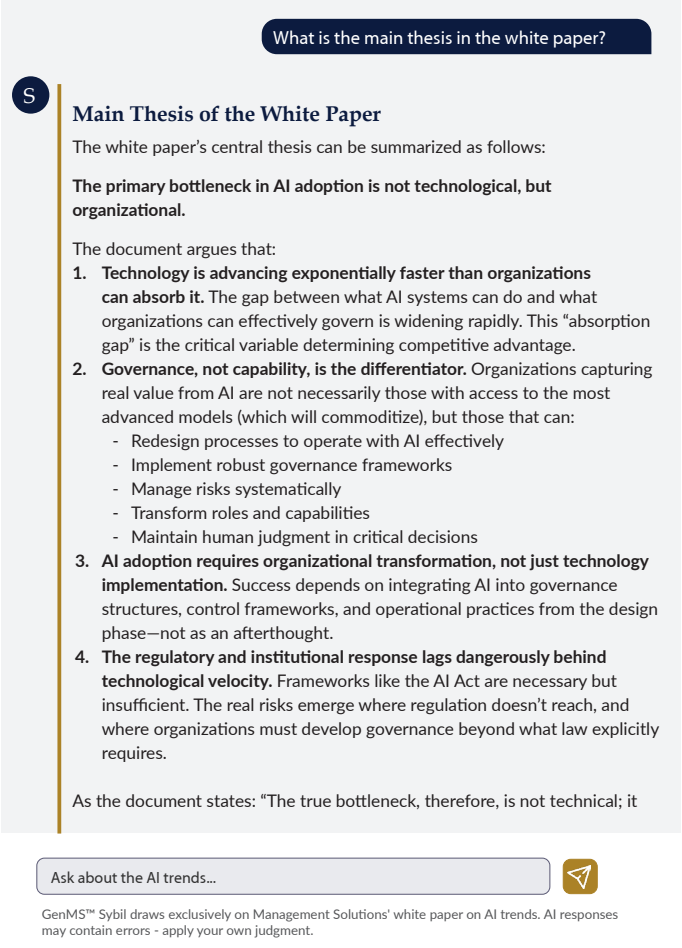
O *prompt* do sistema GenMS™ Sybil tem várias páginas. Ele codifica as proteções comportamentais, os limites operacionais, o tratamento de perguntas fora do escopo e os princípios éticos que regem as respostas do sistema. Seu conteúdo não é publicado na íntegra por motivos de segurança. Sua extensão reflete um princípio que este documento articula na seção de ética: o gap entre os valores declarados e o comportamento real de um sistema de IA é fechada nas instruções concretas que o regem, e não nos princípios que o enquadram.

Tendências em ação

O GenMS™ Sybil não ilustra as tendências discutidas neste documento: ele as implementa. A democratização da IA generativa possibilitou que perfis sem experiência em engenharia de software produzissem um sistema de produção. O *vibe coding* foi o método de construção, não a metáfora. O LLMOps estruturou um processo que, de outra forma, não poderia ser repetido. Os perfis envolvidos combinam conhecimento comercial com a capacidade de executar sistemas cognitivos: o perfil que a análise de talentos identifica como o mais escasso e decisivo. A IA auditou a IA em a fase de segurança. O GenMS™ Atlas aplicou a validação sistemática onde a validação ad hoc teria sido insuficiente. A regulação era um critério de projeto, não um requisito para compliance posterior.

O gargalo que este documento identifica repetidamente como o fator limitante para a adoção real da IA não é tecnológico. Esse caso não mostra que ele é fácil de superar; mostra que ele é superável e que a distância entre descrevê-lo e ultrapassá-lo é, hoje, principalmente uma questão de prontidão organizacional.

Fig. 11. GenMS™ Sybil.



08 | Conclusão

«"A questão não é mais se a IA transformará nosso mundo. A questão é se vamos governar essa transformação juntos ou deixar que ela nos governe».

António Guterres[228]

As tendências discutidas convergem em uma variável central: a velocidade com que o gap entre o que os sistemas de IA podem fazer e o que as organizações podem governar está se acelerando. Este gap – e não o modelo de linguagem, o agente autônomo ou o robô – é o objeto central de gestão dos próximos anos.

O diferencial competitivo sustentável não está no acesso aos modelos mais avançados, que se tornarão progressivamente comoditizados, mas na velocidade e no rigor com que uma organização é capaz de se reprojetar para operar com eles de forma eficaz e responsável. As organizações que estão capturando o valor real têm uma característica em comum: elas entenderam a adoção da IA como uma transformação organizacional, não como um projeto de tecnologia. Governança que habilita em vez de desacelerar, treinamento que transforma em vez de certificar, estruturas de risco que gerenciam a incerteza sem sacrificar a velocidade.

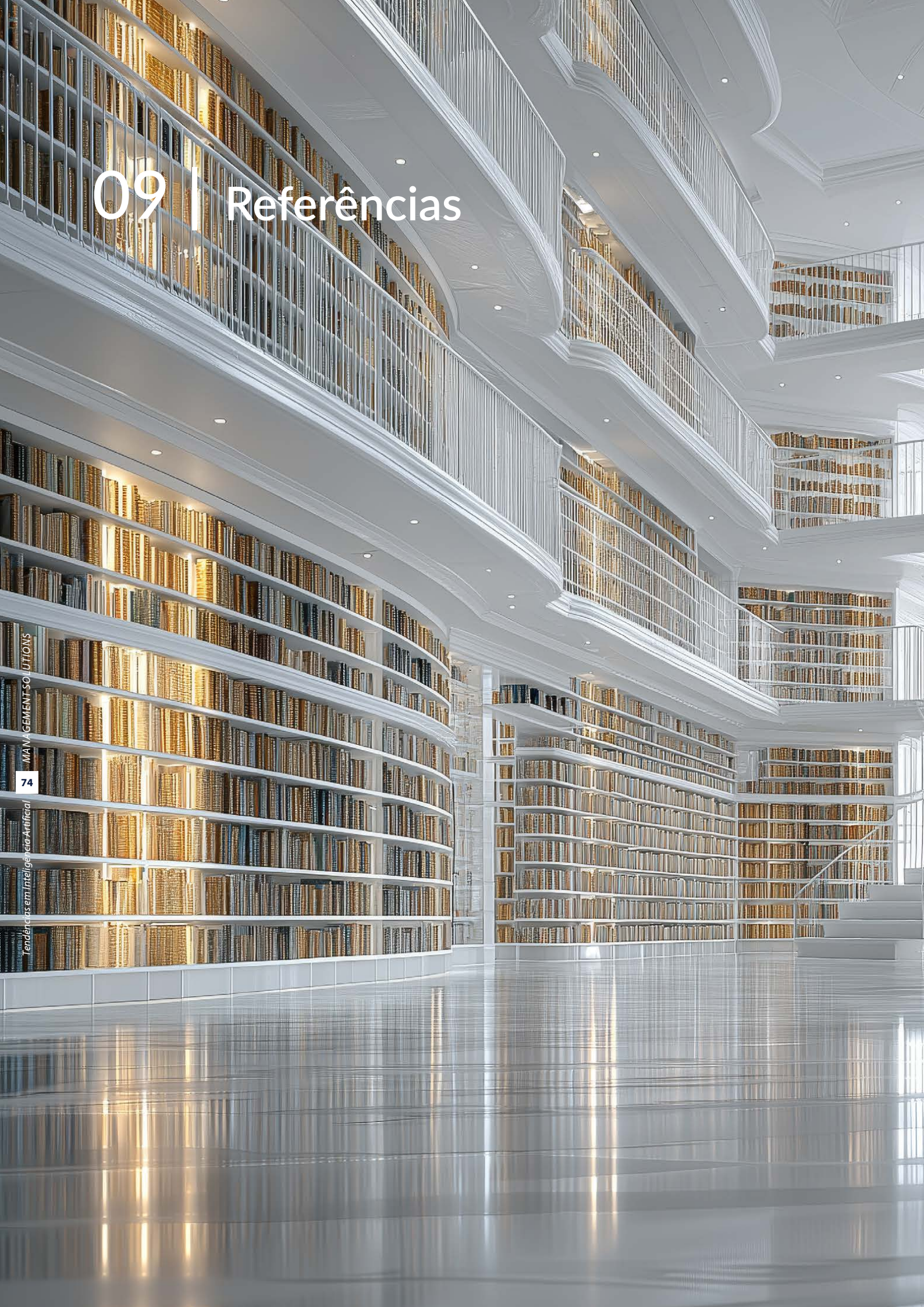
Estruturas regulatórias, padrões técnicos e princípios éticos são necessários, mas não suficientes. A Lei de IA classifica o risco; as normas ISO e NIST estruturam o gerenciamento; os frameworks de ética produzem princípios operacionais. Nenhum deles, por si só, resolve a questão subjacente a várias das tendências analisadas: que tipo de instituição está sendo implantada, que relacionamento ela estabelece com as pessoas que a utilizam e que obrigações isso cria além do que a regulação atual exige explicitamente. É justamente onde a regulação não alcança que os riscos são menos visíveis e têm o maior potencial de causar danos.

O gap entre a curva de capacidade tecnológica e a curva de absorção organizacional está aumentando a cada dia. A tecnologia avança independentemente da velocidade interna de tomada de decisões de cada organização; o ajuste organizacional, por outro lado, depende dela. É essa assimetria que torna a governança da IA uma variável estratégica de primeira ordem, comparável em impacto à própria capacidade técnica.

Os riscos transcendem a competitividade individual. A distribuição dos benefícios da IA, a preservação do julgamento humano nas decisões que o exigem, a capacidade das instituições de manter a legitimidade em sistemas que evoluem mais rapidamente do que as estruturas projetadas para governá-los: essas são dimensões que nenhuma organização pode gerenciar isoladamente e para as quais a resposta institucional está, no momento, significativamente atrasada em relação à velocidade do problema

229 António Guterres (n. 1949), nono secretário-geral das Nações Unidas e ex-primeiro-ministro de Portugal, que tem sido um importante impulsionador da agenda internacional sobre governança de IA e riscos de tecnologias emergentes..

09 | Referências



ABILab (2026). AI Banking (R)evolution: oltre la scelta. Rapporto AI Hub.

AESIA (2026). Agencia Española de Supervisión de Inteligencia Artificial. https://aesia.digital.gob.es/es

AI Act (2024). Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

AI Board (2026). Governance and coordination; AI board meetings. https://digital-strategy.ec.europa.eu/en/policies/ai-board

AICerts (2026). Generative AI Phishing Boosts Clicks, Reshapes Cyber Risk. https://www.aicerts.ai/news/generative-ai-phishing-boosts-clicks-reshapes-cyber-risk/

Altman (2025a). Three Observations. https://blog.samaltman.com/three-observations

Altman (2025b). The Gentle Singularity. https://blog.samaltman.com/the-gentle-singularity

Altman (2024a). Could AI create a one-person unicorn? Fortune. https://finance.yahoo.com/news/could-ai-create-one-person-120000722.html

Altman (2024b). The Intelligence Age. Blog personal. https://ia.samaltman.com/

Amazon (2023). Amazon announces 8 innovations to better deliver for customers, support employees, and give back to communities around the world. https://www.aboutamazon.com/news/operations/amazon-delivering-the-future-2023-announcements

Amodei (2024a). Machines of loving grace. https://www.darioamodei.com/essay/machines-of-loving-grace

Amodei (2024b). Machines of Loving Grace: How AI Could Transform the World for the Better. Blog personal. https://www.darioamodei.com/essay/machines-of-loving-grace

Amodei (2025). Technology in the World, Annual Meeting Davos 2025, World Economic Forum. https://www.weforum.org/meetings/world-economic-forum-annual-meeting-2025/sessions/technology-in-the-world/

Amodei (2026). The Adolescence of Technology. https://www.darioamodei.com/essay/the-adolescence-of-technology

Anthropic (2025). Anthropic Economic Index – September 2025 Report. https://www.anthropic.com/research/anthropic-economic-index-september-2025-report

Anthropic (2026). Claude's new constitution. Anthropic. https://www.anthropic.com/news/claude-new-constitution

Australia (2025). Australia's AI Ethics Principles. https://www.industry.gov.au/publications/australias-ai-ethics-principles

Backlinko (2025). ChatGPT / OpenAI Statistics: How Many People Use ChatGPT? https://backlinko.com/chatgpt-stats

Baker McKenzie (2025). Navigating Labor's Response to AI: Proactive Strategies for Multinational Employers Across the Atlantic. https://www.theemployerreport.com/2025/06/navigating-labors-response-to-ai-proactive-strategies-for-multinational-employers-across-the-atlantic/

Barkhuus (2003). Is Context-Aware Computing Taking Control Away from the User? Three Levels of Interactivity Examined. UbiComp 2003. Springer. https://doi.org/10.1007/978-3-540-39653-6_12

Batty (2024). Digital Twins in City Planning. *Nature Computational Science*, 4, 192–199. https://doi.org/10.1038/s43588-024-00606-5

Bettencourt (2024). Recent Achievements and Conceptual Challenges for Urban Digital Twins. Nature Computational Science, 4, 150–153. https://doi.org/10.1038/s43588-024-00604-7

Bimpas (2024). Leveraging Pervasive Computing for Ambient Intelligence: A Survey on Recent Advancements, Applications and Open Challenges. Computer Networks, 239, 110156. https://doi.org/10.1016/j.comnet.2023.110156

Bletchley Declaration (2023). The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023. https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023

Bloomberg (2026). AI Startup Nabs $100 Million to Help Firms Predict Human Behavior. https://www.bloomberg.com/news/articles/2026-02-12/ai-startup-nabs-100-million-to-help-firms-predict-human-behavior

Boston Dynamics (2025a). An Electric New Era for Atlas. https://bostondynamics.com/blog/electric-new-era-for-atlas/

Boston Dynamics (2025b). Large Behavior Models and Atlas Find New Footing. https://bostondynamics.com/blog/large-behavior-models-atlas-find-new-footing/

Brown (2025). AI's War in the Courtroom: Copyright Disputes Spike in 2025. https://www.bestlawfirms.com/articles/ai-war-in-the-courtroom-copyright-disputes-spike-in-2025/7186

Business Insider (2025a). Walmart just showed off its new AI-powered warehouses — take a look inside. https://www.businessinsider.com/see-inside-walmart-high-tech-refrigerated-grocery-warehouse-2024-7

Business Insider (2025b). The guy who coined '*vibe coding*' predicts it will 'terraform software and alter job descriptions'. https://www.businessinsider.com/andrej-karpathy-coined-vibecoding-ai-prediction-2025-12

Cambridge (2025). Navigating China's regulatory approach to generative artificial intelligence and large language models. https://www.cambridge.org/core/journals/cambridge-forum-on-ai-law-and-governance/article/navigating-chinas-regulatory-approach-to-generative-artificial-intelligence-and-large-language-models/969B2055997BF42DE693B7A1A1B4E8BA

Centre for European Policy (2026). Competition in Generative AI: Updated Assessment. cepInput No. 1/2026. https://www.cep.eu/fileadmin/user_upload/cep.eu/Studien/cepInput_Competition_in_Generative_AI/cepInput_Competition_in_GenAI_Updated_Assessment.pdf

Chatgptiseatingtheworld (2026). Updated Master chart of copyright, DMCA and other claims in suits v. AI (Dec. 5, 2025). https://chatgptiseatingtheworld.com/2025/12/03/updated-master-chart-of-copyright-dmca-and-other-claims-in-suits-v-ai-dec-3-2025/

Chen (2025). Need Help? Designing Proactive AI Assistants for Programming. CHI 2025. ACM. https://doi.org/10.1145/3706598.3714002

Cheong (2025). E2E Process Automation Leveraging Generative AI and IDP-Based Automation Agent: A Case Study on Corporate Expense Processing. https://arxiv.org/abs/2505.20733

Christensen (1997). *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Harvard Business School Press.

CKGSB (2025). Cheung Kong Graduate School of Business. Banking on data: How MYbank is revolutionizing supply chain finance. CKGSB Knowledge. https://english.ckgsb.edu.cn/knowledge/article/unleashing-innovation-in-china-series-banking-on-data-how-mybank-is-revolutionizing-supply-chain-finance/

Corrêa (2023). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *Patterns*, 4(10), 100857. https://doi.org/10.1016/j.patter.2023.100857

Covington (2025). New Artificial Intelligence Legislation in Mexico. Global Policy Watch. https://www.globalpolicywatch.com/2025/03/new-artificial-intelligence-legislation-in-mexico/

CrowdStrike (2025). CrowdStrike Advances Next-Gen SIEM with Threat Hunting Across Data Sources, AI-Driven UEBA. https://www.crowdstrike.com/en-us/blog/crowdstrike-advances-next-gen-siem-capabilities/

Cyberhaven (2025). AI Adoption and Risk Report Q2 2025. https://info.cyberhaven.com/hubfs/Content %20PDF/Cyberhaven %20Labs %20- %202025 %20AI %20Adoption %20& %20Risk %20Report.pdf

Darktrace (2025). New Report Finds that 78 % of Chief Information Security Officers Globally are Seeing a Significant Impact from AI-Powered Cyber Threats – up 5 % from last year. https://www.darktrace.com/news/new-report-finds-that-78-of-chief-information-security-officers-globally-are-seeing-a-significant-impact-from-ai-powered-cyber-threats

DBS Bank (2024). DBS AI-Powered Digital Transformation. https://www.dbs.com/artificial-intelligence-machine-learning/artificial-intelligence/dbs-ai-powered-digital-transformation.html

Dealroom (2025). AI startups: Revenue per employee benchmarks. https://x.com/dealroomco/status/1914264599505018989

Deutsche Bank (2025). Claudio de Sanctis, Head of Private Bank, Deutsche Bank AG Private Bank. Investor Deep Dive 2025. https://investor-relations.db.com/files/documents/other-presentations-and-events/2025/IDD-2025-Script-Private-Bank-Claudio-de-Sanctis.pdf

DHL (2024). DHL Supply Chain Passes Unprecedented 500 Million Picks Milestone Using Locus Robotics Autonomous Mobile Robots. https://www.dhl.com/es-en/home/press/press-archive/2024/dhl-supply-chain-passes-unprecedented-500-million-picks-milestone-using-locus-robotics-autonomous-mobile-robots.html

EBA (2021). EBA Discussion Paper on *Machine Learning* for IRB Models. https://www.eba.europa.eu/sites/default/files/document_library/Publications/Discussions/2022/Discussion %20on %20machine %20learning %20for %20IRB %20models/1023883/Discussion %20paper %20on %20machine %20learning %20for %20IRB %20models.pdf

ECB (2025). ECB Guide to Internal Models. https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.supervisory_guide202507.en.pdf

EDPB (2025). AI Privacy Risks & Mitigations Large Language Models (LLMs). https://www.edpb.europa.eu/our-work-tools/our-documents/support-pool-experts-projects/ai-privacy-risks-mitigations-large_en

Epoch (2025a). AI Benchmarking. https://epoch.ai/benchmarks

Epoch (2025b). How much power will frontier AI training demand in 2030? https://epoch.ai/blog/power-demands-of-frontier-ai-training

ESOMAR (2024). Global Market Research 2024. https://shop.esomar.org/knowledge-center/library?publication=3019

Eurostat (2025). 32.7 % of EU people used generative AI tools in 2025. https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20251216-3

Financial News London (2025). Deutsche Bank to roll out 'banking butlers' for affluent clients. https://www.fnlondon.com/articles/deutsche-bank-to-roll-out-banking-butlers-for-ultra-wealthy-clients-77e0349a

Figure (2025). F.02 Contributed to the Production of 30,000 Cars at BMW. https://www.figure.ai/news/production-at-bmw

FirstPageSage (2025). ChatGPT Usage Statistics: December 2025. https://firstpagesage.com/seo-blog/chatgpt-usage-statistics/

Fortune (2025a). Deloitte allegedly cited AI-generated research in a million-dollar report for a Canadian provincial government. https://fortune.com/2025/11/25/deloitte-caught-fabricated-ai-generated-research-million-dollar-report-canada-government/

Fortune (2025b). Elon Musk reveals massive plans for Tesla and Optimus—'Things are really going to go ballistic next year'. https://fortune.com/2025/01/30/elon-musk-reveals-massive-plans-tesla-optimus-self-driving-cars-humanoid-robots/

Fortune (2025c). AI enabled Klarna to halve its workforce—now, the CEO is warning other tech leaders to be honest about the risks. https://fortune.com/2025/10/10/klarna-ceo-sebastian-siemiatkowski-halved-workforce-says-tech-ceos-sugarcoating-ai-impact-on-jobs-mass-unemployment-warning/

Gartner (2025a). Hype Cycle for Artificial Intelligence. https://www.gartner.com/en/newsroom/press-releases/2025-08-05-gartner-hype-cycle-identifies-top-ai-innovations-in-2025

Gartner (2025b). Gartner Predicts Over 40 % of Agentic AI Projects Will Be Canceled by End of 2027. https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027

Google (2023). Practitioners Guide to MLOps: A framework for continuous delivery and automation of *machine learning*. https://cloud.google.com/resources/mlops-whitepaper

Google DeepMind (2025). AlphaFold: Science and impact. https://deepmind.google/science/alphafold/

Google Quantum AI (2024). Quantum error correction below the surface code threshold. *Nature*, 638, 920–926. https://doi.org/10.1038/s41586-024-08449-y

Grieves-Vickers (2017). Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. En Kahlen, F.J. et al. (eds.), *Transdisciplinary Perspectives on Complex Systems*. Springer. https://doi.org/10.1007/978-3-319-38756-7_4

Gu (2025). Large Language Models for Constructing and Optimizing *Machine Learning* Workflows: A Survey. https://arxiv.org/html/2411.10478v1

Heydari (2025). Tiny *Machine Learning* and On-Device Inference: A Survey of Applications, Challenges, and Future Directions. Sensors, 25(10), 3191. https://doi.org/10.3390/s25103191

HelpNetSecurity (2025). 67 % of daily security alerts overwhelm SOC analysts. https://www.helpnetsecurity.com/2023/07/20/soc-analysts-tools-effectiveness/

Hernández-Torres (2025). Challenges and Opportunities of Ambient Intelligence (AmI) in the 21st Century: A Historical Review. Evolutionary Intelligence, 18, 80. https://doi.org/10.1007/s12065-025-01010-z

Huang (2021). Power of data in quantum *machine learning*. *Nature Communications*, 12, 2631. https://doi.org/10.1038/s41467-021-22539-9

Hyundai (2025). Hyundai Motor Group to Unveil AI Robotics Strategy at CES 2026. https://www.hyundai.com/worldwide/en/newsroom/detail/0000001093

IBM (2025). Cost of a Data Breach Report 2025. https://www.ibm.com/reports/data-breach

IBM (2025b). Chief AI Officers cut through complexity to create new paths to value. https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/chief-ai-officer

Index Ventures (2026). Life, the Universe, and Simile: Leading Simile's $100M Series A. https://www.indexventures.com/perspectives/life-the-universe-and-simile-leading-similes-series-a/

iDanae (3Q20). Cátedra iDanae (UPM–Management Solutions). MLOps, a key element in the digital ecosystem. 2020. https://blogs.upm.es/catedra-idanae/wp-content/uploads/sites/698/2020/10/Idanae-3Q20.pdf

iDanae (2Q23). Cátedra iDanae (UPM–Management Solutions). Large Language Models: a new era for Artificial Intelligence. 2023. https://blogs.upm.es/catedra-idanae/wp-content/uploads/sites/698/2023/07/Idanae-2Q23.pdf**iDanae (1Q24).** Cátedra iDanae (UPM–Management Solutions). Towards a sustainable Artificial Intelligence. 2024. https://blogs.upm.es/catedra-idanae/wp-content/uploads/sites/698/2024/04/Idanae-1Q24-VDef.pdf

iDanae (1Q25). Cátedra iDanae (UPM–Management Solutions). The challenge of biases in the construction of Artificial Intelligence systems. 2025. https://blogs.upm.es/catedra-idanae/wp-content/uploads/sites/698/2025/04/Idanae-1Q25.pdf

iDanae (2Q25). Cátedra iDanae (UPM–Management Solutions). GenAI: an approach to multi-agents systems. 2025. https://blogs.upm.es/catedra-idanae/wp-content/uploads/sites/698/2025/07/Idanae-2Q25.pdf

IEA (2025a). Electricity 2025: Analysis and forecast to 2027. International Energy Agency. https://www.iea.org/reports/electricity-2025

IEA (2025b). World Energy Outlook Special Report on Energy and AI. International Energy Agency. https://www.iea.org/reports/energy-and-ai

IMD (2023). In the Field with Ping An. *IMD Business School*. https://www.imd.org/research-knowledge/digital/articles/in-the-field-with-ping-an/

IMF (2024). Gen-AI: Artificial Intelligence and the Future of Work. https://www.imf.org/-/media/files/publications/sdn/2024/english/sdnea2024001.pdf

ILO (2025a). International Labour Organization. Generative AI and Jobs: A Global Analysis of Potential Effects on Job Quantity and Quality. https://www.ilo.org/sites/default/files/2025-05/WP140_web.pdf

ILO (2025b). International Labour Organization. Governing AI in the World of Work: A review of global ethics guidelines. https://www.ilo.org/resource/article/governing-ai-world-work-review-global-ethics-guidelines

ILO (2025c). International Labour Organization. Global Case Studies of Social Dialogue on AI and Algorithmic Management. https://www.ilo.org/sites/default/files/2025-07/wp144_web.pdf

Inter-Parliamentary Union (2025). Parliamentary actions on AI policy. https://www.ipu.org/impact/democracy-and-strong-parliaments/artificial-intelligence/parliamentary-actions-ai-policy

Ironscales (2025). Ironscales Fall 2025 Threat Report. https://ironscales.com/hubfs/Landing %20Page %20Assets/Fall %202025 %20Threat %20Report/2025 %20Fall %20Threat %20Report_Beyond %20Detection_Reality %20of %20Deepfake %20Attacks %202.pdf

ISO/IEC (2023). ISO/IEC 42001:2023 - Artificial Intelligence Management Systems. https://www.iso.org/standard/42001

Jones (2025). Large Language Models Pass the Turing Test. https://arxiv.org/abs/2503.23674

Jumpcloud (2025). How Effective Is AI for Cybersecurity Teams? 2025 Statistics. https://jumpcloud.com/blog/how-effective-is-ai-for-cybersecurity-teams

KnowBe4 (2025). Phishing Threat Trends Report. https://www.knowbe4.com/hubfs/Phishing-Threat-Trends-2025_Report.pdf

Krijger, J. (2023). Operationalising ethics for AI in the financial industry: Insights from the Volksbank case study. *Journal of Digital Banking*, 8(3), 220–241. https://doi.org/10.69554/YQZC2796

Kumar (2020). Adversarial *Machine Learning* - A Taxonomy and Terminology of Attacks and Mitigations. NIST AI 100-2e2023. https://csrc.nist.gov/pubs/ai/100/2/e2023/final

Kusumegi et al. (2025). Scientific production in the era of large language models. Science. https://www.science.org/doi/10.1126/science.adw3000

Li (2025). Pitfalls and prospects of quantum *machine learning*. *Nature Computational Science*, 5, 1095–1097. https://doi.org/10.1038/s43588-025-00914-6

Liu (2024). Towards provably efficient quantum algorithms for large-scale machine-learning models. *Nature Communications*, 15, 434. https://doi.org/10.1038/s41467-023-43957-x

Lukac (2025). Ambient AI Scribes in Clinical Practice: A Randomized Trial. NEJM AI. https://doi.org/10.1056/AIoa2501000

Management Solutions (2023). Explainable Artificial Intelligence (XAI): Challenges of model interpretability. 2023. https://www.managementsolutions.com/en/microsites/whitepapers/explainable-artificial-intelligence

Management Solutions (3Q23). Follow-up report on *Machine Learning* for IRB models. Technical note on regulations, 3Q23, 2023.

https://www.managementsolutions.com/en/publications-and-events/regulatory-notes/technical-notes-on-regulations/follow-report-machine-learning-irb-models

Management Solutions (4Q24). Artificial Intelligence Act. Technical note on regulations, 4Q24, 2024. https://www.managementsolutions.com/en/publications-and-events/regulatory-notes/technical-notes-on-regulations/proposal-regulation-european-approach-artificial-intelligence

Management Solutions (4Q24a). Artificial Intelligence: regulatory landscape. Technical note on regulations, 4Q24, 2024. https://www.managementsolutions.com/en/publications-and-events/regulatory-notes/technical-notes-on-regulations/artificial-intelligence-regulatory-landscape

Marwala (2025). Why the Need for Governing Ambient Intelligence Has Never Been More Urgent. United Nations University. https://unu.edu/article/why-need-governing-ambient-intelligence-has-never-been-more-urgent

Mascelli (2025). Harvest Now Decrypt Later: Examining Post-Quantum Cryptography and the Data Privacy Risks for Distributed Ledger Networks. *Finance and Economics Discussion Series* 2025-093. Board of Governors of the Federal Reserve System. https://doi.org/10.17016/FEDS.2025.093

Microsoft (2026). AI-powered SIEM, built for modern security. https://marketingassets.microsoft.com/gdc/gdccKuCLQ/original

MIT Technology Review (2024). What's next for AlphaFold: A conversation with a Google DeepMind Nobel laureate. https://www.technologyreview.com/2025/11/24/1128322/whats-next-for-alphafold-a-conversation-with-a-google-deepmind-nobel-laureate/

MITRE (2025). ATLAS (Adversarial Threat Landscape for Artificial-Intelligence Systems). https://atlas.mitre.org/

Moroni (2025). Insurmountable Limitations of City-Scale Digital Twins? On Urban Knowledge and Planning. Computational Urban Science. Springer Nature. https://doi.org/10.1007/s43762-025-00174-0

Nadella (2025). LinkedIn Post. https://www.linkedin.com/posts/satyanadella_just-wrapped-our-earnings-call-and-wanted-activity-7389433821181562880-GnMz/

Nissenbaum (1996). Accountability in a Computerized Society. Science and Engineering Ethics, 2, 25–42. https://link.springer.com/article/10.1007/BF02639315

NIST (2023). AI Risk Management Framework. https://www.nist.gov/itl/ai-risk-management-framework

NIST (2024a). Secure Software Development Practices for Generative AI and Dual-Use Foundation Models: An SSDF Community Profile. https://www.nist.gov/publications/secure-software-development-practices-generative-ai-and-dual-use-foundation-models-ssdf

NIST (2024b). Post-Quantum Cryptography Standards: FIPS 203, FIPS 204, FIPS 205. National Institute of Standards and Technology. https://csrc.nist.gov/projects/post-quantum-cryptography

Notateslaapp (2025). Tesla Eyes $20K Price Target For Optimus, Extremely Fast Production Ramp. https://www.notateslaapp.com/news/3314/tesla-eyes-20k-price-target-for-optimus-extremely-fast-production-ramp

OECD (2024). A Sectoral Taxonomy of AI Intensity. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/12/a-sectoral-taxonomy-of-ai-intensity_c2baae71/1f6377b5-en.pdf

OECD (2025a). Emerging Divides in the Transition to Artificial Intelligence. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/06/emerging-divides-in-the-transition-to-artificial-intelligence_eeb5e120/7376c776-en.pdf

OECD (2025b). The effects of generative AI on productivity, innovation and entrepreneurship. OECD Artificial Intelligence Papers, no. 39. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/06/the-effects-of-generative-ai-on-productivity-innovation-and-entrepreneurship_da1d085d/b21df222-en.pdf

OECD (2026). AI Use by Individuals Surges Across the OECD as Adoption by Firms Continues to Expand. https://www.oecd.org/en/about/news/announcements/2026/01/ai-use-by-individuals-surges-across-the-oecd-as-adoption-by-firms-continues-to-expand.html

Ouyang (2025). FELA: A Multi-Agent Evolutionary System for Feature Engineering of Industrial Event Log Data. https://arxiv.org/html/2510.25223

OWASP (2025). OWASP Top 10 for Large Language Model Applications. https://owasp.org/www-project-top-10-for-large-language-model-applications/

Oxford (2025). AI Regulation: The Politics of Fragmentation and Regulatory Capture. https://blogs.law.ox.ac.uk/oblb/blog-post/2025/06/ai-regulation-politics-fragmentation-and-regulatory-capture

Parfit (1984). Reasons and Persons. Oxford University Press.

Park (2023). Generative Agents: Interactive Simulacra of Human Behavior. UIST '23: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology. ACM. https://doi.org/10.1145/3586183.3606763

Park (2024). Generative Agent Simulations of 1,000 People. https://arxiv.org/abs/2411.10109

Patel (2025). U.S. AI Law & Policy Explained. Enterprise AI Governance. https://oliverpatel.substack.com/p/us-ai-law-and-policy-explained

Pew (2025). Pew Research Institute. How the U.S. Public and AI Experts View Artificial Intelligence. https://www.pewresearch.org/

Phishcare (2025). Top 10 Deepfake Phishing Scams. https://phishcare.com/top-10-deepfake-phishing-scams/

Pixiebrix (2025). Top Chief AI Officers of 2025. https://www.pixiebrix.com/reports/top-ai-officers-of-2025

PR Newswire (2023). SlashNext's 2023 State of Phishing Report Reveals a 1,265 % Increase in Phishing Emails Since the Launch of ChatGPT in November 2022, Signaling a New Era of Cybercrime Fueled by Generative AI. https://www.prnewswire.com/news-releases/slashnexts-2023-state-of-phishing-report-reveals-a-1-265-increase-in-phishing-emails-since-the-launch-of-chatgpt-in-november-2022--signaling-a-new-era-of-cybercrime-fueled-by-generative-ai-301971557.html

Proofpoint (2025). AI Threat Detection. https://www.proofpoint.com/us/threat-reference/ai-threat-detection

Pu (2025). Assistance or Disruption? Exploring and Evaluating the Design and Trade-offs of Proactive AI Programming Support. CHI 2025. ACM. https://doi.org/10.1145/3706598.3713357

Quartz (2025). AI powers smaller startups toward a new era of unicorns. https://qz.com/unicorn-entrepreneur-founder-solo-ai-startup-automation-workforce

Rawls (1971). A Theory of Justice. Harvard University Press.

Ryt Bank (2025). The World's First AI-Powered Bank. https://www.rytbank.my/

Sacra (2026). Cursor revenue, valuation & funding. https://sacra.com/c/cursor/

Shan (2024). Transitioning from MLOps to LLMOps: Navigating the Unique Challenges of Large Language Models. https://doi.org/10.3390/info16020087

Shankar (2021). Towards Observability for *Machine Learning* Pipelines. DOI:10.48550/arXiv.2108.13557. (PDF) Towards Observability for *Machine Learning* Pipelines

SoSafe (2025). Global businesses face escalating AI risk, as 87 % hit by AI cyberattacks. https://sosafe-awareness.com/company/press/global-businesses-face-escalating-ai-risk-as-87-hit-by-ai-cyberattacks/

SQ Magazine (2025). AI Cyber Attacks Statistics 2025: How Attacks, Deepfakes & Ransomware Have Escalated. https://sqmagazine.co.uk/ai-cyber-attacks-statistics/

Stanford (2023a). Stanford Encyclopedia of Philosophy. Ethics of Artificial Intelligence and Robotics. https://plato.stanford.edu/entries/ethics-ai/

Stanford (2023b). Stanford Encyclopedia of Philosophy. Philosophy of Artificial Intelligence. https://plato.stanford.edu/entries/artificial-intelligence/

Stanford (2025). The 2025 AI Index Report. https://hai.stanford.edu/ai-index/2025-ai-index-report

Stone (2025). Navigating MLOps: Insights into Maturity, Lifecycle, Tools, and Careers. https://arxiv.org/html/2503.15577v1

Tesla Car World (2025). Elon Musk Unveils Tesla Bot Gen 3 Real Homemaker Updates. https://www.youtube.com/watch?v=HR1HrrneNHs&t=1s

Teslarati (2025). Tesla Optimus' pilot line will already have an incredible annual output. https://www.teslarati.com/tesla-optimus-pilot-line-will-already-have-an-incredible-annual-output/

The Network Installers (2025). AI Cyber Threat Statistics. https://thenetworkinstallers.com/es/blog/ai-cyber-threat-statistics/

Thompson (1980). Moral Responsibility of Public Officials: The Problem of Many Hands. American Political Science Review, 74(4), 905–916. https://www.cambridge.org/core/journals/american-political-science-review/article/abs/moral-responsibility-of-public-officials-the-problem-of-many-hands/39DD3FAB7BF7DC7A242407143674F22B

Toyota Research Institute (2025). AI-Powered Robot by Boston Dynamics and Toyota Research Institute Takes a Key Step Towards General-Purpose Humanoids. https://www.tri.global/news/ai-powered-robot-boston-dynamics-and-toyota-research-institute-takes-key-step-towards-general

UK AI Safety Institute (2026). International AI Safety Report 2026. UK Government. https://www.gov.uk/government/publications/international-ai-safety-report-2026

UK DSIT (2023). UK Department for Science, Innovation and Technology. Public Attitudes to Data and AI: Tracker Survey (Report). https://www.gov.uk/government/publications/public-attitudes-to-data-and-ai-tracker-survey

UK Government (2023). A pro-innovation approach to AI regulation. Policy paper. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper

UNDP (2025). Human Development Report 2025. United Nations Development Programme. https://hdr.undp.org/content/human-development-report-2025

UNESCO (2023). Guidance for Generative AI in Education and Research. https://unesdoc.unesco.org/ark:/48223/pf0000386693

United Nations (2025). The Sustainable Development Goals Report 2025. United Nations, Department of Economic and Social Affairs. https://unstats.un.org/sdgs/report/2025/

WatchGuard (2025). Evasive Malware Surges 40 % in WatchGuard's Latest Internet Security Report. https://www.watchguard.com/es/wgrd-news/blog/evasive-malware-surges-40-watchguards-latest-internet-security-report

WIPO (2025). WIPO Conversation on Intellectual Property and Frontier Technologies. https://www.wipo.int/en/web/frontier-technologies/frontier_conversation

World Bank (2024). Global Trends in AI Governance. https://documents1.worldbank.org/curated/en/099120224205026271/pdf/P1786161ad76ca0ae1ba3b1558ca4ff88ba.pdf

World Bank (2025). World Development Report 2025: Standards for Development. World Bank. https://www.worldbank.org/en/publication/wdr2025

World Economic Forum (2025a). AI in Action: Beyond Experimentation to Transform Industry. https://reports.weforum.org/docs/WEF_AI_in_Action_Beyond_Experimentation_to_Transform_Industry_2025.pdf

World Economic Forum (2025b). Transforming Consumer Industries in the Age of AI. https://reports.weforum.org/docs/WEF_Transforming_Consumer_Industries_in_the_Age_of_AI_2025.pdf

World Health Organization (2024). Ethics and Governance of Artificial Intelligence for Health. https://iris.who.int/server/api/core/bitstreams/f780d926-4ae3-42ce-a6d6-e898a5562621/content

Yuan et al. (2025). The Impact of AI Adoption in the Workplace on Employees: A Systematic Review. https://www.researchgate.net/publication/396219396_The_Impact_of_AI_Adoption_in_the_Workplace_on_Employees_A_Systematic_Review

10 | Glossário

AGI (Artificial General Intelligence): um sistema de IA capaz de realizar qualquer tarefa cognitiva que possa ser realizada por um ser humano, com generalização transferível entre domínios. Um horizonte estratégico debatido cuja definição precisa carece de consenso científico.

AI Act: Regulamento (UE) 2024/1689, a primeira estrutura jurídica abrangente sobre IA. Classifica os sistemas por nível de risco (inaceitável, alto, limitado, limitado, mínimo) e impõe obrigações estruturais aos sistemas de alto risco. Penalidades de até €35 milhões ou 7% do faturamento global.

AI Board: Fórum para coordenação e interpretação comum entre a Comissão Europeia e as autoridades nacionais de supervisão de IA, criado pelo AI Act.

AI Office: órgão técnico central da Comissão Europeia responsável pela supervisão de modelos de IA de uso geral (GPAI) nos termos do AI Act.

AI-enhanced: modelo organizacional no qual a IA é usada para otimizar os processos existentes sem redesenhá-los a partir de recursos de IA. Estágio predominante na maioria das organizações atualmente.

AI-first: modelo organizacional no qual os processos e a estrutura são projetados com base nos recursos de IA, atribuindo o julgamento humano apenas a tarefas em que sua vantagem comparativa é inequívoca.

AI-only: um modelo organizacional hipotético no qual as operações principais dispensam funcionalmente o trabalho humano.

AIMS (AI Management System): sistema de gerenciamento de IA de acordo com a ISO/IEC 42001, equivalente à ISO 27001 para segurança cibernética, mas específico para IA: abrange políticas, avaliações de impacto, controle de fornecedores e monitoramento contínuo.

Alucinação: comportamento intrínseco de modelos generativos por meio do qual eles produzem informações falsas ou imprecisas apresentadas com aparente confiança. Não se trata de um bug ocasional, mas de uma consequência estrutural do projeto atual dos LLMs.

Ambient AI: IA que opera sem ser explicitamente invocada: observa continuamente o contexto, infere necessidades e age proativamente. A interface desaparece; o próprio ambiente se torna o ponto de interação.

Ambient scribe: Sistema de IA ambiente implantado em ambientes clínicos que ouve a conversa entre o médico e o paciente e gera automaticamente a documentação do encontro sem instruções explícitas do usuário.

BEC (Business Email Compromise): um tipo de ataque cibernético no qual a identidade de um gerente ou fornecedor é personificada para desviar fundos ou extrair informações. A IA generativa possibilita isso por meio de *deepfakes* de áudio e vídeo altamente confiáveis.

Lacuna de absorção: distância cada vez maior entre a curva de capacidade da tecnologia de IA (exponencial, autoacelerada) e a curva de absorção organizacional (redesenho de processos, reengenharia de funções, governança).

CAIO / CDAIO (Chief AI Officer / Chief Data and AI Officer): função executiva responsável pela liderança estratégica da IA em uma organização.

Citizen data scientist: profissional não técnico capaz de realizar análises básicas de dados com ferramentas visuais. A IA generativa vai além desse conceito, fornecendo recursos analíticos avançados diretamente para usuários finais não técnicos.

Computação quântica: paradigma computacional que explora as propriedades quânticas (superposição, emaranhamento) para resolver determinados problemas intratáveis para sistemas clássicos.

Criptografia resistente a ataques quânticos (PQC): conjunto de algoritmos criptográficos projetados para resistir a ataques de computadores quânticos. O NIST publicou os primeiros padrões em 2024 (FIPS 203, 204, 205).

Dark LLM: modelos de linguagem modificados especificamente para o crime cibernético (WormGPT, FraudGPT, GhostGPT). Eles geram malware, explorações e campanhas de engenharia social sem restrições éticas. Comercializados na dark web com suporte técnico.

Data poisoning: ataque adversário que consiste em injetar dados maliciosos no conjunto de treinamento de um modelo para degradar seu comportamento ou introduzir vieses controlados pelo invasor.

Deepfake: conteúdo audiovisual sintético gerado por IA que imita a aparência ou a voz de uma pessoa real. Usado em ataques de BEC, fraude e desinformação com uma taxa de sucesso significativamente maior do que o phishing tradicional.

Differential privacy: técnica que adiciona ruído estatístico controlado aos dados ou resultados para impedir a identificação de indivíduos, preservando a utilidade estatística agregada.

DPIA (Data Protection Impact Assessment): avaliação do impacto da proteção de dados exigida pelo GDPR (Art. 35). O EDPB a considera obrigatória na maioria das implantações de LLM, dado seu processamento sistêmico de dados pessoais.

Edge AI / TinyML: capacidade de executar modelos de IA diretamente em dispositivos como smartphones, wearables e sensores sem depender de conectividade com a nuvem. Principal facilitador da IA ambiente.

Efeito Bruxelas: fenômeno pelo qual a regulação da UE (AI Act, GDPR) força as empresas globais a adaptarem seus produtos aos padrões europeus devido ao volume de mercado, exportando de fato essa regulação para o resto do mundo.

Ethics washing: fenômeno em que uma organização publica princípios éticos de IA sem traduzi-los em controles operacionais, responsabilidades concretas ou mecanismos de auditoria.

Evasão adversária: um ataque que introduz perturbações imperceptíveis nas entradas de um modelo para causar classificações ou respostas incorretas durante a inferência, sem modificar o próprio modelo.

Explicabilidade: capacidade de descrever o funcionamento interno de um modelo de IA de forma que seja compreensível para diferentes públicos (regulador, cliente, funcionário). Requisito técnico em modelos regulados; implementável em ML com técnicas como SHAP, LIME ou análise de sensibilidade.

Feature engineering: processo de criação de variáveis preditivas a partir de dados brutos para alimentar modelos de AM. Uma das fases mais intensivas em conhecimento especializado do ciclo de vida clássico de AM; significativamente acelerada pela IA generativa.

Federated learning: paradigma de treinamento distribuído no qual os dados permanecem em dispositivos locais e somente as atualizações de modelos são compartilhadas. Atenua os riscos de privacidade em LLMs ao custo de maior complexidade operacional.

GDPR (General Data Protection Regulation): Regulamento de Proteção de Dados (UE) 2016/679. Seus princípios de minimização, direito de ser esquecido e transparência apresentam tensões estruturais com a arquitetura e o ciclo de vida dos LLMs.

Gêmeo Digital (Digital Twin): uma simulação dinâmica de um sistema físico ou humano que é atualizada em tempo real com dados do sistema real. Funciona com alta confiabilidade em sistemas físicos determinísticos; os LLMs abriram sua extensão para o comportamento humano e sistemas sociais complexos.

GPAI (General Purpose AI): modelo de IA de propósito geral (por exemplo, GPT, Claude, Gemini) capaz de executar uma ampla variedade de tarefas.

Gradient boosting: técnica de ML que combina várias árvores de decisão sequencialmente para melhorar a previsão.

Harvest now, decrypt later: estratégia em que agentes estatais capturam comunicações criptografadas hoje com a intenção de descriptografá-las quando a computação quântica amadurecer. Ameaça presente aos dados com uma longa vida útil de confidencialidade.

Hub & spokes: modelo organizacional de IA com um Centro de Excelência central que estabelece recursos transversais, enquanto equipes descentralizadas em linhas de negócios desenvolvem soluções específicas com relatórios multifuncionais para o hub.

IA agêntica: sistemas de IA que planejam, executam tarefas complexas e operam de forma autônoma em infraestruturas corporativas reais, além de responder a avisos. Recursos incrementais: estado persistente, planejamento dinâmico, execução em sistemas reais e orquestração de vários agentes.

IA generativa: família de modelos capazes de gerar conteúdo original (texto, imagens, áudio, vídeo, código) em resposta a instruções de linguagem natural. Baseada em arquiteturas de transformadores em larga escala.

IRB (Internal Ratings-Based): abordagem regulatória que permite que os bancos usem modelos internos para calcular o capital regulatório para o risco de crédito.

Jagged intelligence: conceito de Andrej Karpathy para descrever o perfil de habilidades dos LLMs atuais: brilhantes em tarefas complexas (olimpíadas de matemática) e frágeis em tarefas aparentemente simples (comparação de números decimais).

Jailbreak: uma técnica para contornar os controles de segurança de um modelo de IA por meio de instruções projetadas para fazer com que o sistema ignore suas restrições comportamentais.

LIME (Local Interpretable Model-agnostic Explanations): técnica de explicabilidade que gera aproximações locais de um modelo complexo para explicar previsões individuais.

LLM (Large Language Model): modelo de linguagem em grande escala baseado na arquitetura do transformador, treinado em vastos corpora textuais. Base técnica de parte da IA generativa atual. Exemplos: GPT, Claude, Gemini, Llama.

LLMOps (Large Language Model Operations): extensão específica do MLOps para gerenciar as propriedades dos LLMs em produção: comportamento não determinístico, avisos como superfície de risco, alucinações, custo por token e rastreabilidade das interações.

Lock-in de fornecedor: dependência estrutural de um único modelo ou fornecedor de infraestrutura que dificulta a migração (reescrever integrações, recertificação de compliance, custos proibitivos). Risco estratégico na adoção de IA.

LoRA (Low-Rank Adaptation): técnica eficiente de ajuste fino de LLM que adapta o modelo básico a um domínio específico modificando apenas um subconjunto de parâmetros, reduzindo drasticamente os recursos computacionais necessários.

Machine learning (ML): ramo da IA que desenvolve algoritmos capazes de aprender padrões a partir de dados históricos sem serem explicitamente programados para cada tarefa. Diferentemente da IA generativa, os modelos clássicos de ML classificam, preveem e otimizam; eles não geram conteúdo.

Machine unlearning: conjunto de técnicas experimentais que buscam remover seletivamente o conhecimento derivado de dados específicos de um modelo já treinado, em resposta ao direito do GDPR de ser esquecido.

Malware polimórfico: código malicioso que altera sua assinatura para evitar a detecção. A IA generativa é responsável por isso: variantes avançadas reescrevem seu código a cada 15 segundos, mantendo a funcionalidade idêntica, e derrotam sistemas baseados em assinaturas estáticas.

Many hands problem: problema de responsabilidade difusa em sistemas complexos em que o dano ocorre sem intenção deliberada e a cadeia causal é fragmentada entre vários atores (designers, treinadores, implantadores, usuários).

MCP (Model Context Protocol): padrão aberto que padroniza como os modelos de IA interagem com aplicativos, fontes de dados e ferramentas externas. Ele elimina o débito técnico de integrações proprietárias e torna cada ferramenta um ativo reutilizável para qualquer agente.

MLOps (*Machine Learning* **Operations)**: um conjunto de processos padronizados e recursos tecnológicos para criar, implantar e operacionalizar modelos de ML de forma confiável durante todo o seu ciclo de vida. Ele abrange a preparação de dados, a experimentação, a validação, a implementação e o monitoramento.

Model drift: degradação silenciosa do desempenho de um modelo de produção causada por alterações na distribuição dos dados de entrada em relação aos dados de treinamento.

Multimodalidade: a capacidade de um modelo de IA de processar e gerar simultaneamente vários tipos de informações (texto, imagens, áudio, vídeo, código) em uma única arquitetura de conversação.

NIST AI RMF: estrutura de gerenciamento de riscos de IA do Instituto Nacional de Padrões e Tecnologia. Organizada nas funções

Governar, Mapear, Medir e Gerenciar. Focado em recursos de IA confiáveis.

Orquestração multiagente: Arquitetura na qual vários agentes de IA especializados colaboram com um coordenador central para atingir objetivos complexos, gerenciando dependências, prioridades e transferências de informações entre agentes.

Phishing hiperpersonalizado: ataques de *phishing* gerados por IA que analisam perfis públicos e estilo de redação corporativa para criar mensagens altamente personalizadas.

Prompt: uma instrução ou entrada de linguagem natural que um usuário fornece a um sistema de IA generativo para obter uma resposta.

Prompt injection: ataque no qual instruções maliciosas incorporadas em entradas (documentos, URLs, dados externos) manipulam o comportamento do modelo para ignorar restrições ou executar ações não autorizadas.

RAG (Retrieval-Augmented Generation): arquitetura que complementa um LLM com um sistema de recuperação de informações externas em tempo real. Ela reduz as alucinações ancorando as respostas em documentos verificáveis e separa o conhecimento atualizável da memória estática do modelo.

ReAct (Reason + Act): loop de controle de agente de IA que combina raciocínio (análise de situação) e ação (uso de ferramentas ou APIs) iterativamente. A base do núcleo cognitivo dos sistemas agênticos.

Reskilling: o processo de aquisição de novas competências para desempenhar funções profissionais diferentes das atuais, em resposta à transformação do trabalho pela IA. Alavanca estratégica devido ao desequilíbrio estrutural entre a oferta e a demanda de talentos em IA.

Robótica humanoide: robôs com forma e capacidade de movimento semelhantes aos humanos, integrados a modelos de IA para percepção, raciocínio e aprendizado.

Shadow AI: uso não autorizado de ferramentas de IA generativas em ambientes corporativos, geralmente em plataformas públicas sem garantias de privacidade.

SHAP (SHapley Additive exPlanations): técnica de explicabilidade baseada na teoria dos jogos que quantifica a contribuição de cada variável para uma previsão individual de um modelo.

SIEM / XDR / SOAR: plataformas de segurança cibernética: SIEM (Security Information and Event Management), XDR (Extended Detection and Response) e SOAR (Security Orchestration, Automation and Response).

Soberania tecnológica: a capacidade de um estado ou organização de controlar camadas essenciais da cadeia de valor da IA (hardware, infraestrutura, modelos, talentos) para manter a autonomia estratégica.

Technoblocks: esferas parcialmente incompatíveis de influência tecnológica (lideradas pelos EUA, China e Europa) com suas próprias lógicas de governança, segurança e valores.

Teste de Turing: Critério proposto por Alan Turing (1950) para avaliar se uma máquina apresenta comportamento inteligente indistinguível da conversação humana.

Transformer: arquitetura de rede neural escalável baseada em mecanismos de atenção, introduzida pelo Google em 2017. Base técnica de todos os LLMs modernos.

UEBA (User and Entity Behaviour Analytics): sistemas de segurança cibernética que estabelecem linhas de base dinâmicas de comportamento normal e detectam anomalias.

Upskilling: o processo de expansão das competências existentes para se adaptar aos novos requisitos da mesma função profissional, especificamente no contexto da adoção de IA.

Vibe coding: um paradigma de desenvolvimento de software por meio de conversas iterativas em linguagem natural com sistemas de IA que interpretam requisitos, geram aplicativos completos, detectam erros e produzem testes e documentação automaticamente. Termo cunhado por Andrej Karpathy.

Nosso objetivo é superar as expectativas dos nossos clientes tornando-nos parceiros de confiança

A Management Solutions é uma firma internacional de serviços de consultoria especializada em consultoria empresarial, de riscos, organizacional e de processos, tanto em seus componentes funcionais quanto na implementação das tecnologias relacionadas.

Com sua equipe multidisciplinar (funcionais, matemáticos, técnicos, etc.) de mais de 4.000 profissionais, a Management Solutions opera por meio de seus 52 escritórios (22 na Europa, 24 nas Américas, 3 na Ásia, 1 na África e 2 na Oceania).

Para atender às necessidades de seus clientes, a Management Solutions estruturou suas práticas por setores (Instituições Financeiras, Energia, Telecomunicações e outros setores) e por linhas de atividade, abrangendo uma ampla gama de competências – Estratégia, Gestão Comercial e Marketing, Gestão e Controle de Riscos, Informação Gerencial e Financeira, Transformação: Organização, Processos e Tecnologia, e Novas Tecnologias e Metodologias.

Javier Calvo
Sócio e Chief AI Officer da Management Solutions
javier.calvo.martin@managementsolutions.com

Manuel Ángel Guzmán
Sócio da Management Solutions
manuel.guzman@managementsolutions.com

Segismundo Jiménez
Diretor da Management Solutions
segismundo.jimenez@managementsolutions.com

Louis Perron
Gerente da Management Solutions
louis.perron@managementsolutions.com

Management Solutions, serviços profissionais de consultoria

Management Solutions é uma firma internacional de serviços de consultoria focada na assessoria de negócio, riscos, finanças, organização e processos.

Para mais informações acesse **www.managementsolutions.com**

Siga-nos em:

www.managementsolutions.com

Madrid Barcelona Bilbao Coruña Málaga London Frankfurt Düsseldorf Wien Paris Bruxelles Amsterdam Copenhagen Oslo Stockholm Warszawa Wrocław Zürich Milano Roma Bologna Lisboa Beijing Abu Dhabi İstanbul Johannesburg Sydney Melbourne Toronto New York New Jersey Boston Pittsburgh Columbus Atlanta Birmingham Houston Phoenix Miami SJ de Puerto Rico San José Ciudad de México Monterrey Querétaro Medellín Bogotá Quito São Paulo Rio de Janeiro Lima Santiago de Chile Buenos Aires