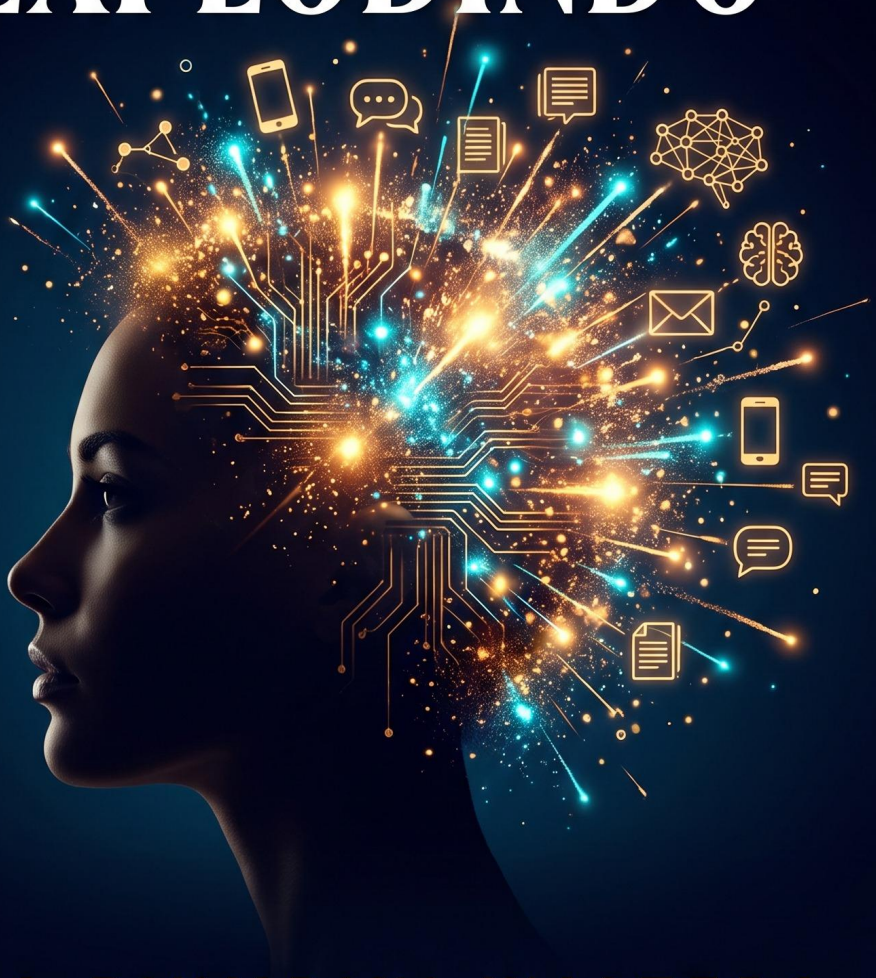
MINHA CABEÇA ESTÁ EXPLODINDO



ANA PIUMBINI WALTERS

Minha Cabeça Está Explodindo

Como a inteligência artificial está mudando o trabalho, as gerações, a comunicação e o futuro da humanidade

ANA PIUMBINI WALTERS

Edição completa — 10 de outubro de 2026

SUMÁRIO

Prólogo – Eu reconheço este filme, mas o final é diferente.....4

Parte I – Acorde: o mundo já mudou.....11

Capítulo 1 – Minha cabeça está explodindo.....12

Capítulo 2 – O mundo começou a mudar toda semana.....25

Capítulo 3 – Depois do Google.....39

Capítulo 4 – Velocidade sem direção.....51

Parte II – O que é o quê.....61

Capítulo 5 – Empresa, laboratório, modelo, produto e agente....62

Capítulo 6 – Open source, open weight, fechado e frontier.....74

Capítulo 7 – Como comparar inteligências sem ser enganado....86

Capítulo 8 – O Atlas Vivo das Novas Mentes.....98

Parte III – A corrida pela fronteira.....110

Capítulo 9 – OpenAI, Anthropic e Google: três ideias de futuro.111

Capítulo 10 – O Google está ficando para trás?.....124

Capítulo 11 – A China não é apenas DeepSeek.....134

Capítulo 12 – Quatro estratégias nacionais para a mesma revolução.....146

Capítulo 13 – Musk, X, Grok e SpaceX.....158

Capítulo 14 – De IBM e Google aos frontier labs.....170

Capítulo 15 – IPOs, valuations e a corrida por capital.....180

Capítulo 16 – Nokia, BlackBerry e o momento Apple da IA.....190

Parte IV – A compressão dos negócios.....203

Capítulo 17 – De zero a US$100 milhões em meses.....204

Capítulo 18 – A empresa de uma pessoa.....214

Capítulo 19 – Quando o preço da inteligência cai.....226

Capítulo 20 – Quem souber competir, competirá.....236

Parte V – Aprender, trabalhar e contratar de novo.....247
Capítulo 21 – A faculdade ainda importa?.....248
Capítulo 22 – PhD na era da inteligência abundante.....259
Capítulo 23 – A escola que ensina para um mundo que terminou 269
Capítulo 24 – Como contratar quando todos podem parecer excelentes no papel.....281
Capítulo 25 – Prompt não é magia.....292

Parte VI – Quatro gerações na mesma revolução.....302
Capítulo 26 – Quinze, vinte, vinte e cinco, trinta.....303
Capítulo 27 – Quarenta, cinquenta, sessenta, setenta.....312
Capítulo 28 – Uma ligação já é o segundo nível de intimidade. 324
Capítulo 29 – Quando a IA entra na amizade, no amor e na identidade.....334

Parte VII – O que ainda depende de nós.....347
Capítulo 30 – As três sociedades da era da IA.....348
Capítulo 31 – Inteligência não é sabedoria.....359

Epílogo – A escolha não é usar ou não usar.....370
Apêndice A – Plano de 30 dias para acordar.....375
Apêndice B – O Teste Ana de Inteligência Artificial.....377
Apêndice C – Laboratório de Prompts – Método CLARO.....379
Apêndice D – Glossário de Duas Gerações.....385

PRÓLOGO

Eu reconheço este filme, mas o final é diferente

Eu estava assistindo ao podcast Moonshots, do Peter Diamandis, no YouTube, quando ouvi uma frase que fez meu cérebro parar por alguns segundos. A China, disseram, havia colocado milhares de inteligências artificiais dentro de uma simulação para formar uma sociedade. Agentes com perfis, memórias, crenças e comportamentos interagindo dentro de um mundo digital.

Minha reação não foi elegante nem científica. Foi imediata:

Minha cabeça está explodindo.

Quando esta conversa começou, eu disse: "Tenho 55 anos e minha cabeça está explodindo." Não porque eu nunca tivesse visto uma tecnologia transformadora. Exatamente pelo contrário. Eu reconhecia o sentimento. Já o havia vivido outras vezes.

Eu não assisti às grandes transições tecnológicas apenas da arquibancada. Na Microsoft Brasil, trabalhei com a linha do Microsoft Office quando ainda era necessário explicar por que uma família ou uma empresa precisaria de um computador pessoal. No varejo, participei da entrada dos computadores nos supermercados, tirando aquela máquina de um território reservado a especialistas e colocando-a diante de consumidores comuns. Na IBM Brasil, tornei-me a primeira mulher Diretora Nacional de Vendas da linha PSG, de computadores pessoais, para

o varejo, e trabalhei com a linha Aptiva. Mais tarde, acompanhei Nokia e Motorola em mais de 1.500 pontos de venda. Vi o computador deixar de ser objeto de técnicos. Vi o telefone deixar de ser telefone. Vi a internet deixar de ser um destino e se tornar o ambiente onde a vida acontece.

Cada uma dessas mudanças parecia enorme enquanto estava acontecendo. E era. O computador pessoal aumentou nossa capacidade de calcular, escrever, organizar e armazenar. A internet distribuiu informação e conectou mercados. O celular libertou a comunicação de um lugar fixo. O smartphone reuniu câmera, mapa, música, banco, loja, jornal, escritório e identidade em um objeto que carregamos no corpo.

Mas aquilo que eu estava vendo agora parecia diferente.

O computador sempre havia executado instruções. Agora ele conversava. Ele podia ler um contrato, analisar um desenho, comparar fornecedores, escrever um e-mail delicado, programar um sistema, criar uma imagem, traduzir uma reunião, ensinar um idioma, pesquisar um mercado e sugerir uma estratégia. E estava começando a fazer algo ainda mais importante: receber objetivos, escolher etapas, usar ferramentas e agir.

Eu precisava saber se a história do podcast era verdadeira. A primeira descoberta já continha uma lição para este livro. A realidade era menos cinematográfica do que a manchete e, ao mesmo tempo, mais perturbadora de uma forma intelectual.

O trabalho chamado Light Society descrevia uma estrutura capaz de simular sociedades com mais de um bilhão de agentes. Isso não significava um bilhão de seres conscientes, nem um bilhão de AGIs pensando livremente dentro de um computador. Em sua escala máxima, o sistema dependia de operações simplificadas e componentes substitutos para tornar a simulação

possível. Ainda assim, o salto era extraordinário: pesquisadores estavam construindo laboratórios computacionais para observar confiança, opiniões, informação e comportamento emergindo em populações artificiais de escala planetária.[1]

Menos Hollywood. Mais consequência.

A diferença importa. Este livro não pode repetir cada manchete impressionante sem examinar o que existe por baixo dela. Mas também não pode usar a imperfeição de uma manchete como desculpa para ignorar o acontecimento. O que estamos construindo não precisa ser consciente para mudar a sociedade. Não precisa ser AGI para alterar empregos, relações, negócios, educação, guerra, política e poder.

Enquanto eu tentava compreender aquilo, pensei em pessoas de todas as idades. Pensei em alguém de 70 anos que ainda acredita que tecnologia é uma coisa para os filhos resolverem. Pensei em alguém de 50 que administra uma empresa, conhece profundamente seu setor, mas sente vergonha de não entender a nova linguagem. Pensei em alguém de 30 que acredita ter muito tempo. Pensei em alguém de 20 que usa a IA para estudar, trabalhar e conversar, mas talvez ainda não tenha percebido que a ferramenta também está ajudando a formar sua identidade. Pensei em uma criança de 15 anos que nunca precisará lembrar de um mundo no qual respostas não conversavam de volta.

E pensei na mudança que parece pequena até olharmos de perto. Durante décadas, pesquisar significou abrir o Google, digitar palavras, receber uma lista de páginas e decidir em qual link confiar. Agora, uma geração inteira começa pela pergunta direta a uma inteligência. Em vez de procurar documentos, ela solicita síntese. Em vez de navegar, conversa. Em vez de encontrar apenas informação, pede explicação, comparação, decisão e ação.

Isso economiza um tempo extraordinário. Também muda a relação com a verdade. Quando alguém abre cinco páginas, ainda enxerga a existência de cinco fontes. Quando recebe uma resposta perfeita em segundos, pode esquecer que aquela frase foi construída por um sistema que interpreta, seleciona e, às vezes, inventa.

A mesma transformação está acontecendo na comunicação. Antigamente, você pegava o telefone e ligava. Hoje, para muitas pessoas, uma ligação já é o segundo nível de intimidade. Primeiro vem o texto. A mensagem respeita o tempo do outro, permite editar o tom e evita a exposição do improviso. Uma chamada inesperada pode significar urgência, invasão ou proximidade. Em breve, a inteligência artificial poderá escrever o texto de um lado, resumir do outro, suavizar a resposta, traduzir o sentimento e negociar um horário. Duas pessoas poderão manter uma relação enquanto suas inteligências conversam entre si.

Se a IA escreveu meu pedido de desculpas, ele continua sendo meu? Se ela encontrou as palavras que eu não conseguia encontrar, ela tornou a mensagem menos autêntica ou me ajudou a ser mais verdadeiro? Se o e-mail do chefe foi escrito por uma IA e a resposta do funcionário também, houve uma comunicação humana? Essas não são perguntas sobre software. São perguntas sobre humanidade.

Também pensei nas oportunidades. Pela primeira vez, uma pessoa pode ter ao redor de si capacidades que antes exigiam uma empresa inteira. Pesquisa, escrita, design, programação, marketing, atendimento, análise financeira, tradução e operação podem ser ampliados por sistemas artificiais. Isso não significa que todos ficarão ricos. Não significa que uma ideia ruim se tornará boa porque ganhou uma equipe de agentes. Significa que a

distância entre uma pessoa e aquilo que ela consegue tentar ficou menor.

Startups de IA já atravessam em meses marcos de receita que antes consumiam anos. O custo de utilizar determinada capacidade está caindo enquanto o custo de construir a fronteira sobe para bilhões. É a mesma sensação da telefonia em uma escala diferente: primeiro, acesso caro e limitado; depois, presença em todos os lugares. Só que agora não estamos barateando apenas comunicação. Estamos barateando partes do trabalho cognitivo.

Essa democratização pode ampliar pessoas. Pode também aumentar desigualdades. Haverá quem aprenda a dirigir a inteligência artificial, quem seja dirigido por sistemas que não compreende e quem fique excluído por renda, idioma, medo, educação ou falta de infraestrutura. Não são as pessoas que perdem valor humano. Mas competências, métodos de trabalho e modelos de negócio podem perder relevância econômica. Quem se recusar a aprender corre o risco de ficar preso aos métodos que o mercado deixou para trás.

Por isso, este não será um livro sobre qual chatbot ganhou o ranking desta semana. Nós vamos comparar ChatGPT, Claude, Gemini, Grok, Kimi, DeepSeek, Qwen e outros. Vamos explicar open source, open weight, fechado, frontier, agente, AGI, ARR, valuation e IPO. Vamos analisar a diferença entre Estados Unidos e China, a posição do Google, a atração dos frontier labs, a estratégia de Musk, a compressão das startups, a empresa de uma pessoa, a queda do preço da inteligência e o risco de se perder em tantas inovações.

Mas esses nomes serão parte do mapa, não o destino do livro.

A pergunta maior será sempre a mesma: o que acontece conosco quando inteligência, criação, comunicação e execução deixam de ser capacidades exclusivamente humanas?

Eu não escrevo como cientista da computação. Escrevo como empresária, mãe, imigrante, mulher e testemunha de várias revoluções. Escrevo como alguém que aprendeu que gigantes não desaparecem no dia em que o futuro muda de endereço. Primeiro eles parecem invencíveis. Depois protegem aquilo que os fez grandes. Enquanto isso, outra empresa percebe que o produto não é mais o produto.

Vi isso na IBM. Vi isso com Nokia e BlackBerry. Vi a Apple transformar telefone em ecossistema. Vi o Google se tornar o sonho de uma geração de talentos. Agora vejo muitos desses talentos querendo entrar em um frontier lab ou criar sua própria empresa. Isso não significa que o Google terminou. Significa que ele já não possui sozinho o símbolo do futuro.

Também não escrevo sozinha. Estou conversando com a própria tecnologia sobre a qual este livro trata. A IA me ajuda a pesquisar, organizar, comparar, desafiar frases e localizar lacunas. Mas a responsabilidade pelo que entra nestas páginas é humana. A experiência é minha. O julgamento é meu. A decisão de acreditar, cortar, corrigir ou publicar é minha.

Essa talvez seja a primeira lição da nova era: colaborar não é abdicar.

Não precisamos escolher entre idolatrar a inteligência artificial e rejeitá-la. Precisamos aprender a enxergar o que ela é, o que ainda não é, quem a controla, em que ela falha e o que desejamos construir com ela.

A escolha não será simplesmente usar ou não usar. A inteligência artificial entrará em produtos, empresas, governos, escolas, hospitais, relações e decisões mesmo quando não estivermos olhando. A escolha real será participar conscientemente ou permitir que outros definam o sistema no qual viveremos.

Bem-vindo ao momento em que a história parou de pedir licença.

Notas

1. Guan, Haoxiang et al. "Modeling Earth-Scale Human-Like Societies with One Billion Agents." arXiv:2506.12078, jun. 2025 (versão revisada consultada em 6 out. 2026). https://arxiv.org/abs/2506.12078

PARTE I

Acorde: o mundo já mudou

CAPÍTULO 1

Minha cabeça está explodindo

Tese do capítulo: o primeiro passo não é dominar a tecnologia; é compreender emocionalmente que estamos diante de uma mudança de escala histórica.

Eu conheço bem a sensação de ouvir uma notícia e sentir vontade de contá-la imediatamente para alguém. Quem passou a vida trabalhando com vendas, varejo e gente aprende que uma boa história se espalha sozinha. Naquela noite, diante da televisão, eu tinha nas mãos uma história perfeita: a China teria colocado milhares de inteligências artificiais dentro de uma simulação para formar uma sociedade inteira.

Era o tipo de frase que viaja rápido. Cabe numa mensagem de WhatsApp, rende assunto no jantar e assusta na medida certa.

Mas eu também aprendi, em mais de três décadas entre empresas, lançamentos e consumidores, que a primeira versão de uma notícia raramente é a versão completa. O mundo da tecnologia está cheio de produtos anunciados como revolução que desapareceram em poucos meses — e de outros, tratados como modismo, que acabaram mudando tudo.

Então, antes de repetir a manchete, fiz o que este livro vai pedir que você faça muitas vezes: fui conferir.

Não fui conferir porque desconfiava de tudo. Fui conferir porque queria entender o tamanho verdadeiro daquilo. E foi exatamente aí que minha cabeça explodiu pela segunda vez. A primeira explosão foi de espanto. A segunda foi de escala.

Este capítulo é sobre essas duas explosões. Não vou começar explicando como funcionam as redes neurais, nem listando qual modelo ganhou o ranking da semana. Isso virá depois, com calma. O primeiro passo é outro, e é mais humano: aceitar, com o coração e não apenas com a cabeça, que estamos diante de uma mudança de escala histórica.

1. A noite em que a escala deixou de ser abstrata

Existem tecnologias que chegam devagar. Primeiro aparecem em laboratórios, depois em empresas especializadas, depois em produtos caros, e finalmente entram na vida cotidiana. Durante um tempo, quem não presta atenção consegue continuar vivendo quase da mesma maneira.

A inteligência artificial generativa não está seguindo apenas esse caminho. Ela chegou ao público como produto de conversa e, em poucos anos, começou a se espalhar por pesquisa, programação, design, educação, atendimento, medicina, defesa, finanças, robótica e ciência. O intervalo entre demonstração e uso real encolheu. O intervalo entre uma versão e a próxima também.

Por isso, o sentimento dominante não é apenas curiosidade. É desorientação.

Quando eu disse que minha cabeça estava explodindo, eu não estava afirmando compreender tudo. Eu estava reconhecendo que a escala havia ultrapassado a capacidade normal de absorção. Cada semana parecia trazer um modelo mais forte, uma ferramenta mais barata, um agente mais independente, uma startup crescendo em velocidade improvável, um país anunciando outra estratégia, um novo medo e uma nova promessa.

Escala é uma palavra perigosa, porque parece técnica. Quando alguém diz "um bilhão", o cérebro registra apenas "muito". Então deixe-me traduzir do jeito que a gente aprende a traduzir números no varejo, para que eles caibam na vida real. Se você contasse em voz alta um número por segundo, sem dormir, sem comer e sem parar no fim de semana, levaria mais de trinta e um anos para chegar a um bilhão. Um bilhão é quase cinco vezes a população do Brasil. E era dessa ordem de grandeza a sociedade artificial de que eu tinha acabado de ouvir falar.

Naquela noite, a escala deixou de ser abstrata para mim. Não porque eu tivesse entendido a engenharia. Mas porque percebi que já não estávamos falando apenas de uma ferramenta que ajuda uma pessoa a escrever um e-mail. Estávamos falando de sistemas capazes de representar populações inteiras.

O cérebro humano gosta de narrativas com começo, meio e fim. A revolução da IA chega como uma história na qual o capítulo seguinte é publicado antes de terminarmos o anterior.

Isso produz dois erros opostos. O primeiro é transformar cada lançamento em singularidade. O segundo é cansar de tanto exagero e concluir que nada está realmente mudando. Precisamos evitar ambos.

Nem toda demonstração se tornará produto. Nem toda startup sobreviverá. Nem todo benchmark significa superioridade prática. Nem todo agente executa o que promete. Nem toda previsão de substituição acontecerá no prazo anunciado.

Mas o conjunto dos sinais é impossível de ignorar. Modelos que escrevem, programam, percebem imagens, entendem áudio, usam computadores e executam sequências de trabalho estão melhorando. Empresas estão investindo valores extraordinários em compute, energia e talentos. Governos tratam IA como

infraestrutura econômica e questão de segurança nacional. Jovens já a usam para aprender, procurar informação, se divertir e pedir aconselhamento.

Em pesquisa do Pew Research Center publicada em fevereiro de 2026, 57% dos adolescentes americanos disseram já ter usado chatbots para procurar informação, 54% para trabalhos escolares, 16% para conversas casuais e 12% para apoio emocional ou aconselhamento.[1]

O número mais importante talvez não seja o maior. Doze por cento significa que uma tecnologia que mal existia na vida pública começou a ocupar uma posição antes reservada a amigos, pais, professores, líderes religiosos, psicólogos ou outras pessoas de confiança.

A IA não está apenas entrando no trabalho. Está entrando na formação da identidade humana.

2. Um bilhão de agentes não é um bilhão de pessoas

A história da sociedade artificial é um exemplo perfeito de como devemos pensar durante esta revolução.

O título do artigo científico é "Modeling Earth-Scale Human-Like Societies with One Billion Agents".[2] O trabalho apresenta Light Society, uma estrutura para simular interações sociais em grande escala. O resumo afirma que ela suporta sociedades com mais de um bilhão de agentes e demonstra experimentos envolvendo confiança e propagação de opiniões.

A palavra agente pode nos enganar. No cotidiano, imaginamos uma entidade autônoma com memória, objetivos, personalidade e capacidade de decisão. Em um sistema de pesquisa, "agente" pode

representar desde um processo relativamente complexo conduzido por um modelo de linguagem até uma unidade simplificada, cujas decisões são aproximadas por um modelo menor ou reaproveitadas de respostas anteriores.

Na escala de um bilhão, o sistema não estava mantendo um bilhão de conversas conscientes. A engenharia utilizava simplificações e modelos substitutos para representar interações sem chamar um grande modelo a cada instante. Isso reduz dramaticamente o significado da manchete "um bilhão de AGIs". Também revela uma conquista diferente: a capacidade de aproximar comportamentos aprendidos por modelos e testá-los em uma população sintética gigantesca.

Quando terminei de ler, percebi algo curioso. A manchete estava errada nos detalhes e certa no instinto. Não existiam milhares de AGIs vivendo dentro de um computador. Mas existia, sim, uma nova capacidade de construir sociedades de laboratório em uma escala que, até muito pouco tempo atrás, parecia ficção. Para ter uma ideia, em 2023 um dos trabalhos mais comentados da área, feito por pesquisadores de Stanford e do Google, colocou 25 agentes para viver em uma pequena cidade virtual.[3] No fim de 2024, o projeto OASIS já falava em simular redes sociais com até um milhão de agentes.[4] E os próprios autores do Light Society registram que a maioria das estruturas anteriores trabalhava com grupos relativamente pequenos, muitas vezes com menos de mil agentes.[2] Não são exatamente o mesmo tipo de agente, mas a curva impressiona: de vinte e cinco para um bilhão em pouco mais de dois anos.

Essa é a diferença entre ser enganado e ser informado. A pessoa enganada repete "um bilhão de AGIs". A pessoa informada entende o que de fato foi construído — e continua atenta, só que pelos motivos certos.

Esse tipo de simulação pode ajudar pesquisadores a formular hipóteses sobre confiança, difusão de mensagens, polarização ou resposta a políticas. Mas não transforma a simulação em uma previsão perfeita da sociedade real. Seres humanos possuem corpo, história, cultura, instituições, relações, contradições e consequências que um perfil artificial não reproduz completamente.

Uma sociedade sintética pode produzir uma hipótese. Não pode receber automaticamente o status de humanidade em miniatura.

Ao mesmo tempo, não devemos diminuir o impacto por ela não ser consciente. Mercados financeiros são afetados por algoritmos que não sentem. Redes sociais moldam opinião por sistemas que não compreendem o significado humano do que recomendam. Um modelo não precisa possuir experiência subjetiva para influenciar milhões de experiências subjetivas.

A pergunta prática não é apenas "essa IA pensa como nós?". É também "o que acontece quando instituições começam a agir com base no que ela simula, prevê, classifica ou recomenda?".

Imagine uma campanha política testando milhares de mensagens em sociedades sintéticas antes de apresentá-las a pessoas reais. Imagine uma empresa simulando reações de consumidores. Imagine um governo avaliando o efeito de uma política. Imagine um grupo mal-intencionado procurando o ponto mais eficiente para espalhar desconfiança.

Alguns desses cenários serão exagerados. Outros já estão tecnicamente próximos. A nossa obrigação é aprender a distinguir capacidade real, marketing, hipótese e risco.

NÃO CONFUNDA

Um agente de simulação não é necessariamente um chatbot completo. Um chatbot não é automaticamente um agente. Um agente não é automaticamente uma AGI. E nenhuma dessas palavras prova consciência.

3. Por que este momento é diferente

Toda geração acredita estar vivendo uma ruptura única. Por isso, comparar IA com eletricidade, imprensa, internet ou smartphone pode soar como marketing. A comparação só é útil quando mostra sem esconder.

O computador pessoal democratizou poder de cálculo e produção de documentos. A internet democratizou distribuição de informação. O smartphone democratizou acesso contínuo, comunicação e sensores. A inteligência artificial começa a democratizar capacidades de interpretação, síntese, criação e execução.

Eu vivi de perto as três primeiras. Quando trabalhei com o Microsoft Office, quando ajudei a levar computadores para os supermercados e, depois, com a linha Aptiva na IBM, boa parte do trabalho consistia em responder a uma pergunta que hoje parece estranha: "Mas para que serve isso?". O computador precisava ser explicado. Precisava de demonstração, de paciência e de alguém convencendo uma família de que aquela máquina na mesa da sala valia o investimento. Eu fui gerente de produtos da Microsoft no Brasil, cuidando da linha do Microsoft Office, e fui beta tester de muitos softwares da Microsoft e, na época, também da Lotus e da Autodesk. Depois, como diretora nacional de compras do Grupo Pão de Açúcar, ajudei a levar o computador pessoal para as redes

de supermercados brasileiras, negociando contratos internacionais de fornecimento com Compaq, HP, Microsoft, Intel, AMD e IBM. Era tirar aquela máquina do território dos especialistas e colocá-la diante do consumidor comum.

Com o celular, a pergunta já era outra. Quando acompanhei Nokia e Motorola em mais de 1.500 pontos de venda, e também quando fui beta tester do BlackBerry, a conversa girava cada vez mais em torno de qual aparelho, qual plano e quanto custava — e cada vez menos em torno de por que alguém precisaria de um.

Com a inteligência artificial, a pergunta mudou de novo. Para muita gente, ela já não é "para que serve?". É "o que isso ainda não consegue fazer?". E, quando a pergunta muda desse jeito, é sinal de que não estamos diante de mais um aparelho, mas de uma nova camada da vida.

A diferença está na flexibilidade. Uma calculadora resolve uma classe definida de problemas. Um mecanismo de busca encontra páginas. Um aplicativo tradicional executa funções programadas. Um modelo geral pode receber instruções em linguagem comum e adaptar sua saída a milhares de tarefas.

Essa flexibilidade torna a IA parecida com um novo tipo de infraestrutura. Não porque seja mágica ou infinita, mas porque outras indústrias começarão a ser construídas sobre ela. Assim como ninguém precisava fundar uma companhia telefônica para criar um negócio no celular, milhões de pessoas poderão utilizar modelos que não treinaram para construir serviços, produtos e empresas.

As analogias com as revoluções anteriores não servem para provar que "já vimos isso antes, então vai ficar tudo bem". Servem para reconhecer padrões. Nas transições que vivi, alguns movimentos se repetiram: primeiro a desconfiança, depois o

encantamento, depois a pressa, e por fim uma nova normalidade em que ninguém lembra direito como era antes. Também se repetiu um padrão mais duro: empresas e profissionais excelentes nas regras antigas demoraram a perceber que as regras tinham mudado. Voltarei a essa história com calma quando falarmos de IBM, Nokia e BlackBerry.

Mas as analogias também falham, e é importante dizer onde. O computador pessoal esperava em cima da mesa até alguém ligá-lo. A internet esperava que alguém digitasse um endereço. O smartphone esperava um toque na tela. Todas essas tecnologias ampliavam aquilo que nós fazíamos. A inteligência artificial começa a fazer parte do trabalho: lê, resume, sugere, escreve, escolhe etapas e, cada vez mais, age. Pela primeira vez, a ferramenta não amplia apenas a mão, o olho ou a voz. Ela amplia — e, em algumas tarefas, substitui — partes do raciocínio.

A outra diferença é a velocidade de chegada. Pesquisadores do Federal Reserve Bank de St. Louis, da Universidade Vanderbilt e de Harvard compararam essas ondas nos Estados Unidos. Em agosto de 2024, menos de dois anos depois do lançamento do ChatGPT, 39% dos americanos de 18 a 64 anos já usavam IA generativa. Três anos depois do lançamento do IBM PC, em 1981, a adoção do computador pessoal estava em torno de 20%; a da internet, dois anos depois de sua abertura comercial, em 1995, também.[5] No trabalho, o ritmo da IA foi parecido com o do PC; fora dele, foi muito mais rápido. Para quem, como eu, ajudou a levar o computador a famílias que ainda precisavam ser convencidas de sua utilidade, esses números dizem algo que eu já sentia na pele: desta vez, o intervalo entre a novidade e o uso diário encolheu.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: durante muito tempo, o computador foi como uma calculadora gigante: fazia exatamente o que mandavam, nada além. Hoje ele se parece mais com um colega de grupo que lê o material, dá opinião e se oferece para fazer uma parte do trabalho. Rápido, útil — e às vezes errado com muita confiança.

Para quem tem 70 anos: lembre-se de quando o telefone chegou em casa e, depois, de quando foi parar no bolso. A IA é um passo do mesmo tamanho, só que em outra direção: em vez de levar sua voz mais longe, ela começa a ajudar a pensar, escrever e resolver. Ninguém precisou entender a central telefônica para fazer uma ligação; você também não precisa entender a engenharia da IA para começar a usá-la com bom senso.

Ao mesmo tempo, existe um paradoxo. A capacidade fica mais acessível na ponta, enquanto a fronteira fica mais concentrada. Poucas organizações conseguem financiar os chips, a energia, os data centers, os pesquisadores e os dados necessários para treinar os maiores sistemas. Muitas pessoas poderão alugar a inteligência; poucas controlarão sua origem.

Essa tensão entre democratização e concentração atravessará todo este livro.

Ela aparece quando um empreendedor individual ganha uma equipe de agentes e quando uma plataforma decide quais agentes ele pode usar. Aparece quando um estudante recebe um tutor gratuito e quando o tutor influencia o que ele acredita. Aparece quando um pequeno país pode operar um modelo open weight e quando esse país ainda depende de chips fabricados em outro lugar. Aparece quando uma empresa economiza tempo e quando o funcionário não entende por que foi avaliado por um sistema.

O potencial da inteligência artificial parece, neste momento, não ter um horizonte visível. Isso não significa que seja literalmente infinito. Significa que ainda não conseguimos enxergar com segurança onde estão seus limites técnicos, econômicos e sociais.

Diante disso, existem três reações comuns: medo, euforia e paralisia.

O medo enxerga apenas substituição e perda. A euforia enxerga apenas riqueza e abundância. A paralisia tenta esperar até que tudo fique claro.

Nenhuma das três é suficiente.

A resposta mais adulta é curiosidade disciplinada: aprender o suficiente para participar, verificar o suficiente para não ser enganado, experimentar o suficiente para descobrir valor e preservar julgamento suficiente para não entregar a direção da própria vida.

O objetivo não é prever perfeitamente o futuro.
É tornar-se capaz de reconhecê-lo enquanto ele está chegando.

O QUE AINDA NÃO SABEMOS

Onde estão os limites técnicos, econômicos e sociais desses sistemas. Se simulações sociais conseguirão antecipar o comportamento real de pessoas com confiabilidade. Se — e quando — surgirá algo que mereça, por consenso, o nome de AGI. E como a sociedade vai dividir os ganhos e os custos dessa nova capacidade.

Naquela noite, eu não entendi tudo. Ainda não entendo, e desconfio de quem diz que entende. Mas entendi o suficiente para

saber que não dava para voltar a dormir como se nada tivesse acontecido.

Minha cabeça continua explodindo. A diferença é que, agora, eu sei que essa sensação não é sinal de fraqueza nem de idade. É o que acontece quando uma pessoa atenta percebe, antes de ter todas as respostas, que a escala do mundo mudou. Você não precisa entender toda a engenharia para reconhecer que a história mudou de direção.

E há um detalhe que torna tudo ainda mais desafiador: esta revolução não chega uma vez só. Ela chega toda semana. É sobre isso que vamos conversar no próximo capítulo.

O que mudou / O que significa / O que fazer agora

O que mudou: o computador deixou de apenas executar instruções e passou a conversar, interpretar, criar e começar a agir. A escala do que pode ser construído — até sociedades sintéticas com mais de um bilhão de agentes — deixou de caber nas nossas referências habituais.

O que significa: uma tecnologia não precisa ser consciente nem ser AGI para transformar trabalho, relações, política e poder. A tarefa de cada um de nós é separar capacidade real, marketing, hipótese e risco, sem cair no medo, na euforia ou na paralisia.

O que fazer agora: troque a reação automática pela curiosidade disciplinada. Na próxima manchete sobre IA que surpreender você, antes de repassá-la, faça três perguntas: o que exatamente foi construído? Quem está afirmando isso? E o que isso *não* significa?

AÇÃO EM DEZ MINUTOS

Escolha uma atividade real da sua semana que consome tempo, exige leitura, escrita, pesquisa, comparação ou organização. Escreva o resultado desejado em uma frase. Essa será a primeira tarefa prática do plano de 30 dias.

Notas

1. Pew Research Center. "How Teens Use and View AI." 24 fev. 2026 (pesquisa com 1.458 adolescentes de 13 a 17 anos, realizada de 25 set. a 9 out. 2025). https://www.pewresearch.org/internet/2026/02/24/how-teens-use-and-view-ai/

2. Guan, Haoxiang et al. "Modeling Earth-Scale Human-Like Societies with One Billion Agents." arXiv:2506.12078, jun. 2025. https://arxiv.org/abs/2506.12078

3. Park, Joon Sung; O'Brien, Joseph C.; Cai, Carrie J.; Morris, Meredith Ringel; Liang, Percy; Bernstein, Michael S. "Generative Agents: Interactive Simulacra of Human Behavior." UIST '23 (ACM), 2023. arXiv:2304.03442. https://arxiv.org/abs/2304.03442

4. Yang, Ziyi et al. "OASIS: Open Agent Social Interaction Simulations with One Million Agents." arXiv:2411.11581, nov. 2024. https://arxiv.org/abs/2411.11581

5. Bick, Alexander; Blandin, Adam; Deming, David J. "The Rapid Adoption of Generative AI." Federal Reserve Bank of St. Louis, Working Paper 2024-027, set. 2024 (rev. fev. 2025); NBER Working Paper 32966. https://www.stlouisfed.org/on-the-economy/2024/sep/rapid-adoption-generative-ai

CAPÍTULO 2

O mundo começou a mudar toda semana

Tese do capítulo: a velocidade da IA não é apenas uma característica tecnológica; ela altera nossa percepção de tempo, competência e segurança.

No fim do capítulo anterior, eu disse que esta revolução não chega uma vez só. Ela chega toda semana. Parece exagero de quem está empolgada. Não é. É quase uma descrição de calendário.

Eu percebi isso de um jeito pouco glamouroso. Mais de uma vez, terminei de aprender uma ferramenta, descobri onde ficavam os comandos, criei meus pequenos atalhos, comecei a me sentir confortável... e descobri que já existia uma versão nova. Às vezes, mais de uma. Às vezes, de outra empresa, prometendo fazer melhor e mais barato aquilo que eu tinha acabado de aprender a fazer. Durante anos, testar versões novas foi literalmente parte do meu trabalho: fui beta tester de muitos softwares da Microsoft e, na época, também da Lotus e da Autodesk. A diferença é que, agora, a versão nova chega antes de eu terminar de aprender a anterior.

A primeira reação é rir. A segunda, se a gente não toma cuidado, é uma sensação silenciosa e corrosiva: a de que estamos sempre atrasados.

Este capítulo é sobre essa sensação. Sobre de onde ela vem, por que ela engana e como podemos conviver com a velocidade sem deixar que ela decida por nós.

1. De ciclos anuais a lançamentos semanais

Durante a maior parte da minha carreira, a tecnologia tinha calendário. Havia o lançamento, depois o treinamento das equipes, depois o material de ponto de venda, depois a campanha, depois a temporada de vendas. Um produto novo tinha tempo de ser explicado, testado, criticado, vendido e, só então, substituído. Quando eu cuidava de todos os lançamentos da Nokia no Brasil, na Visual Presence, cada aparelho novo chegava às lojas pelas mãos de mais de 200 demonstradores espalhados pelo país.

Esse ritmo moldava a nossa noção de competência. Quem dominava a linha daquele ano era, durante aquele ano, uma pessoa atualizada. Havia tempo para aprender e havia tempo para aproveitar o que se aprendeu.

Na inteligência artificial, esse intervalo encolheu de um jeito que eu nunca tinha visto. Basta olhar para algumas semanas do meio de 2026.

Em 8 de julho, a xAI, que àquela altura já fazia parte da SpaceX, lançou o Grok 4.5. Em 9 de julho, a OpenAI lançou a família GPT-5.6,[1] e, no mesmo intervalo de dois dias, a Meta apresentou o Muse Spark 1.1. Em 16 de julho, a Moonshot AI anunciou o Kimi K3.[2] No dia seguinte, a Artificial Analysis, uma organização de avaliação independente, resumiu o período em um título: quatro lançamentos de fronteira em oito dias. Segundo a mesma avaliação, o número de laboratórios com pelo menos um modelo acima de 50 pontos em seu índice de inteligência tinha passado de dois, no início de junho, para seis.[3] E a semana

seguinte não deu descanso: em 21 de julho, o Google DeepMind publicou o model card do Gemini 3.6 Flash[4] e, em 24 de julho, a Anthropic lançou o Claude Opus 5.[5]

E o calendário não parou aí. Em 30 de julho, a OpenAI cortou em 80% o preço do GPT-5.6 Luna e em 20% o do Terra.[1] Em 2 de agosto, o Qwen anunciou o Qwen3.8-Max[6] e a União Europeia começou a aplicar novas partes do AI Act, embora as obrigações para sistemas de alto risco tenham sido adiadas, dias antes, para 2 de dezembro de 2027 e 2 de agosto de 2028 pelo Regulamento (UE) 2026/1744.[7] Antes disso, no mesmo ano, tinham chegado a prévia do DeepSeek V4 e o GLM-5.2.[8,9] Em menos de quatro semanas, pelo menos sete laboratórios lançaram modelos novos, preços caíram e mudou a regra do jogo em um dos maiores mercados do planeta.

E aquilo foi só o ensaio. Quando revisei este capítulo, em outubro, a cena já tinha se repetido, maior. Em setembro de 2026, a OpenAI lançou o GPT-6 Astra, completou a família com o GPT-6 Sol e o GPT-6 Luna em 22 de setembro e, uma semana depois, lançou o GPT-6.1 Sol.[10] A Meta lançou o Muse Spark 1.3 em 2 de setembro;[11] a DeepSeek, o V4.1-Flash, no dia 10;[12] a SpaceXAI, o Grok 4.7, no dia 21.[13] A Anthropic lançou o Claude Fable 5.1, o Opus 5.5 (dia 22) e o Sonnet 5.5 (dia 28).[14] E, em 30 de setembro, o Google anunciou o Gemini 4 Argon.[15] Até o termômetro mudou: a Artificial Analysis lançou duas versões novas do seu índice em quatro dias, porque a fronteira se movia rápido demais.[16]

Repare também em algo que essa cronologia não diz. Ela não diz que cada lançamento foi uma revolução. Alguns foram saltos importantes; outros, ajustes. Ela não diz quem ganhou. E resultados divulgados pela própria empresa não são verdade neutra. O que a cronologia mostra é outra coisa: a fronteira se move em semanas.

E a velocidade não está só nos modelos. Está nos preços. O AI Index 2025, da Universidade Stanford, registrou uma queda de mais de 280 vezes no custo de inferência de um sistema com capacidade próxima à do GPT-3.5 entre novembro de 2022 e outubro de 2024.[17] Por enquanto, basta registrar que, quando o preço cai tão rápido, quem decidiu ontem com base no custo talvez precise decidir de novo amanhã.

FOTOGRAFIA DA FRONTEIRA — DUAS JANELAS DE 2026

Julho–agosto (registro histórico): Grok 4.5, GPT-5.6, Muse Spark 1.1 e Kimi K3 em oito dias (8–16 jul.); Gemini 3.6 Flash e Claude Opus 5 (21 e 24 jul.); Qwen3.8-Max e novas regras do AI Act (2 ago.).

Setembro: GPT-6 Astra, GPT-6 Sol e Luna (22 set.) e GPT-6.1 Sol (29 set.); Muse Spark 1.3 (2 set.); DeepSeek V4.1-Flash (10 set.); Grok 4.7 (21 set.); Claude Fable 5.1, Opus 5.5 (22 set.) e Sonnet 5.5 (28 set.); Gemini 4 Argon (30 set.).

Fotografia atualizada em 6 de outubro de 2026. Fatos vivos: devem ser revisados antes da impressão e acompanhados no Atlas Vivo da IA.

2. A fadiga de novidade e a sensação permanente de atraso

Quando a novidade chega uma vez por ano, ela é uma festa. Quando chega toda semana, ela vira ruído de fundo. E ruído de fundo cansa.

Eu chamo isso de fadiga de novidade. É aquele momento em que você vê mais uma manchete dizendo que "tudo mudou" e, em vez de curiosidade, sente vontade de fechar o telefone. Não porque

deixou de se importar, mas porque não aguenta mais ter que se importar com tudo ao mesmo tempo.

E não é só impressão minha. Em uma pesquisa encomendada pela Upwork, uma plataforma de trabalho freelance, com 2.500 executivos, funcionários e freelancers dos Estados Unidos, do Reino Unido, da Austrália e do Canadá, realizada entre abril e maio de 2024, 38% dos funcionários disseram se sentir sobrecarregados por ter de usar IA no trabalho. Entre os que já usavam essas ferramentas, 77% disseram que elas tinham, de pelo menos uma forma, reduzido sua produtividade e aumentado sua carga de trabalho, por exemplo com mais tempo gasto revisando o que a IA produz ou aprendendo a usá-la.[18] É um levantamento feito por uma empresa interessada no tema, e deve ser lido como sinal, não como lei. Mas o sinal é claro: a ferramenta que promete aliviar pode, sem direção, virar mais um peso.

A fadiga tem um primo ainda mais perigoso: a sensação permanente de atraso. Ela funciona assim. Você abre as redes sociais e vê alguém dizendo que montou uma empresa com agentes em um fim de semana. Outra pessoa mostra uma ferramenta que você nunca ouviu falar. Um terceiro garante que quem não usa determinado modelo "já ficou para trás". Ao final de dez minutos, você não aprendeu nada, mas tem certeza de que está atrasado.

Quero dizer com todas as letras uma coisa que me ajudou muito: essa sensação mente.

Ela mente porque compara a sua vida inteira com o melhor momento de outras pessoas. Ela mente porque mistura demonstração com uso real, anúncio com adoção, e promessa com resultado. E ela mente, principalmente, porque parte de uma premissa impossível: a de que alguém, em algum lugar, está

acompanhando tudo. Ninguém está. Nem os especialistas. A fronteira se move rápido demais para que uma única pessoa conheça cada ferramenta, cada versão e cada preço.

Mas a velocidade não mexe só com a nossa atenção. Ela mexe com três coisas mais profundas.

A primeira é a nossa percepção de **tempo**. Antes, aprender algo novo era um investimento com prazo longo de retorno. Agora, parece que o conhecimento vem com data de validade impressa na embalagem. Isso faz muita gente se perguntar: "Para que vou aprender, se daqui a três meses muda tudo?". A pergunta é compreensível. A conclusão, como veremos, está errada.

A segunda é a nossa percepção de **competência**. Uma pessoa pode ter vinte, trinta anos de excelência em sua área e, de repente, sentir-se iniciante diante de uma tela. É importante separar as coisas: você não ficou menos competente porque a ferramenta mudou. O seu julgamento, a sua experiência, o seu conhecimento do cliente, do mercado e das pessoas continuam lá. O que mudou foi a interface entre você e o trabalho. E interface se aprende.

A terceira é a nossa percepção de **segurança**. Durante muito tempo, uma boa formação, um bom emprego e uma boa reputação pareciam proteção suficiente para uma carreira. Quando o mundo muda toda semana, essa sensação de chão firme balança. Para quem tem 50 ou 60 anos, ela aparece como medo de ficar obsoleto. Para quem tem 20, aparece como dúvida sobre qual profissão ainda existirá. É a mesma insegurança, vestida com roupas diferentes.

NÃO CONFUNDA

Novidade não é progresso. Lançamento não é adoção. Resultado divulgado pela empresa não é avaliação independente. E estar desatualizado sobre o nome da ferramenta da semana não é o mesmo que estar despreparado para o futuro.

Não estou dizendo que podemos ignorar a velocidade. A velocidade é real. O que precisamos é de outra relação com ela. No Capítulo 4, vou propor uma regra prática para isso, porque tentar acompanhar tudo é uma das formas mais eficientes de não transformar nada. Aqui, quero começar por algo anterior: aprender a olhar para o que muda sem confundir o retrato com a paisagem.

3. Como registrar uma fotografia datada sem confundi-la com uma verdade eterna

Escrever um livro sobre inteligência artificial é um desafio curioso. Quando você estiver lendo estas páginas, parte do que está aqui terá mudado. Algum modelo citado terá ganhado uma versão nova. Algum preço terá caído. Alguma empresa que parecia liderar terá sido ultrapassada, e talvez outra que ninguém comentava esteja no centro da conversa.

Eu poderia fingir que isso não acontece. Preferi assumir o problema e transformá-lo em método.

Por isso, ao longo do livro, você vai encontrar quadros chamados "Fotografia da fronteira". Eles fazem exatamente o que o nome diz: registram como as coisas estavam em uma data específica. A fotografia factual deste livro, por exemplo, foi tirada em 12 de agosto de 2026 e atualizada em 6 de outubro. Entre uma

data e outra, quase todos os nomes do topo mudaram. Uma fotografia não é mentira porque o tempo passou. A foto do seu aniversário de dez anos continua verdadeira. Ela só não descreve quem você é hoje.

O erro não está em fotografar. Está em pendurar a foto na parede e acreditar que ela é uma janela.

Essa distinção vale muito além deste livro. Vale para a reportagem que você leu, para o vídeo que recebeu no grupo, para o ranking que um colega compartilhou e para a opinião que você formou seis meses atrás. Na era da mudança semanal, toda informação sobre IA deveria vir acompanhada de algumas perguntas simples:

1. **Quando?** Qual é a data dessa informação? Ela ainda é atual?
2. **Qual versão?** De qual modelo, produto ou plano estamos falando, exatamente?
3. **Quem afirma?** É a própria empresa, um avaliador independente, um jornalista, um usuário entusiasmado?
4. **Que tipo de afirmação é essa?** Um fato estável, um fato que pode mudar em semanas, uma interpretação ou uma previsão?
5. **Para qual horizonte?** Se for previsão, ela fala dos próximos dois anos, de dois a cinco anos, ou de cinco a dez?

Essas perguntas são a versão de bolso do sistema que usamos para escrever este livro. Nos bastidores, cada informação recebe uma etiqueta: o que é fato estável, o que é fato vivo e precisa ser revisto, o que é inferência, o que é cenário e o que é experiência minha. Os fatos vivos são revisados antes do fechamento do manuscrito, revisados de novo antes da impressão e, depois disso, continuam sendo atualizados no Atlas Vivo da IA, a página digital que acompanha este livro.

Você não precisa criar um sistema tão formal na sua vida. Mas pode adotar o mesmo espírito. Quando ouvir que "o modelo X é o melhor", acrescente mentalmente: "segundo quem, em qual tarefa, em qual data?". Quando ler que "a IA vai substituir tal profissão", pergunte: "em quanto tempo, com que grau de confiança, e o que exatamente seria substituído?". Essas perguntas não deixam você mais lento; deixam você mais difícil de enganar.

Existe ainda um ganho menos óbvio. Quando passamos a ver cada informação como uma fotografia datada, a ansiedade diminui. Você deixa de tentar memorizar a paisagem inteira, que muda o tempo todo, e passa a aprender a ler fotografias. E isso, sim, é uma habilidade que não vence na semana que vem.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: pense em um jogo online que recebe atualização toda semana. Quem decora a "meta" de hoje vai ter que decorar de novo amanhã. Quem entende como o jogo funciona se adapta a qualquer atualização.

Para quem tem 65 anos: lembre-se do jornal de papel. A notícia era verdadeira no dia em que foi impressa, mas ninguém usava o jornal da semana passada para saber a cotação de hoje. Com a IA, é a mesma coisa, só que o jornal envelhece muito mais rápido.

Isso me leva à lição mais importante deste capítulo. Durante muito tempo, ser competente significava saber. Saber a resposta, conhecer o produto, dominar o sistema. Esse tipo de conhecimento continua valioso, mas tem prazo de validade cada vez mais curto. O que não vence é outra coisa: a capacidade de aprender de novo, com calma, sempre que necessário.

Eu já reaprendi muitas vezes na vida. Reaprendi quando o computador entrou nas casas, quando o celular virou outra coisa,

quando imigrei, quando passei a liderar uma indústria nos Estados Unidos. Cada uma dessas vezes me ensinou que o desconforto do começo não é um sinal de que você não consegue. É o preço de entrada de qualquer coisa que vale a pena.

Em 2006, eu me mudei para a Flórida com o meu filho, o Filipe, e com o meu primeiro marido, o Chico. Não viemos começar nada glamoroso: compramos uma pequena empresa de granito em Port Charlotte e começamos a difícil jornada para conseguir os vistos de trabalho. Em 2010, o Chico faleceu de repente. Fiquei num país estrangeiro, à frente de um negócio, com o meu filho e com uma dor que eu não sabia onde colocar. Durante dois anos, lutei contra o luto e a depressão. Reconstruí a vida peça por peça, muitas vezes do jeito mais cauteloso possível, só para atravessar cada dia. Foi o reaprendizado mais difícil de todos: reaprender a viver. E foi ali que eu entendi que uma mão dolorosa não é o fim do torneio.

A diferença, agora, é a frequência. Antes, reaprendíamos de tempos em tempos. Agora, reaprender virou parte da rotina. E, quando algo vira rotina, precisa de método, não de pânico.

O QUE AINDA NÃO SABEMOS

Se este ritmo de lançamentos vai continuar, acelerar ou desacelerar. Quanto de cada novidade se transformará em ganho real no trabalho das pessoas. E quanto tempo empresas, escolas e governos levarão para se adaptar a uma tecnologia que muda mais rápido do que suas próprias regras.

O que mudou / O que significa / O que fazer agora

O que mudou: a tecnologia deixou de seguir ciclos anuais. Na inteligência artificial, modelos, preços e regras mudam em semanas, como mostram as janelas de julho e de setembro de 2026 registradas neste capítulo.

O que significa: a velocidade altera a forma como sentimos o tempo, avaliamos nossa própria competência e enxergamos nossa segurança. A sensação de atraso permanente é compreensível, mas engana: ninguém acompanha tudo, e informação sobre IA é sempre uma fotografia datada, não uma verdade eterna.

O que fazer agora: pare de tentar memorizar a paisagem e aprenda a ler fotografias. Diante de qualquer novidade sobre IA, pergunte: quando, qual versão, quem afirma, que tipo de afirmação e para qual horizonte. A competência central não será memorizar novidades, mas aprender a reaprender sem entrar em pânico.

AÇÃO EM DEZ MINUTOS

Abra uma nota no celular ou um caderno e crie o seu "Diário de Fronteira". Registre a data de hoje, a ferramenta de IA que você usa (ou quer usar), a versão, se souber, e uma frase sobre o que ela faz por você. Uma vez por mês, volte a essa nota e pergunte: o que mudou, e isso muda alguma coisa no meu trabalho real?

Notas

1. OpenAI. "GPT-5.6: Frontier intelligence that scales with your ambition." 9 jul. 2026; atualização de preço em 30 jul. 2026. https://openai.com/index/gpt-5-6/

2. Moonshot AI. Lançamento do Kimi K3 como serviço em 16 jul. 2026; pesos publicados em 27 jul. 2026. https://huggingface.co/moonshotai/Kimi-K3; Unite.AI, "Moonshot Opens Kimi K3 Weights Under a Revenue-Tiered License." 27 jul. 2026. https://www.unite.ai/moonshot-opens-kimi-k3-weights-under-a-revenue-tiered-license/

3. Artificial Analysis. "Four frontier launches in eight days: six labs now field a model above 50 on the Artificial Analysis Intelligence Index." 17 jul. 2026 (lançamentos analisados: Grok 4.5, GPT-5.6, Muse Spark 1.1 e Kimi K3). https://artificialanalysis.ai/articles/four-frontier-launches-in-eight-days-six-labs-now-field-a-model-above-50-on-the-artificial-analysis-intelligence-index

4. Google DeepMind. "Gemini 3.6 Flash – Model Card." 21 jul. 2026. https://deepmind.google/models/model-cards/gemini-3-6-flash/

5. Anthropic. "Introducing Claude Opus 5." 24 jul. 2026. https://www.anthropic.com/news/claude-opus-5

6. Qwen. "Qwen3.8-Max: A New Bar for Coding and Cowork." 2 ago. 2026. https://qwen.ai/blog?id=qwen3.8

7. European Commission. "AI Act – governance and enforcement." Aplicação e fiscalização a partir de 2 ago. 2026. https://digital-strategy.ec.europa.eu/en/policies/ai-act-governance-and-enforcement; Regulamento (UE) 2026/1744 do Parlamento Europeu e do Conselho, de 8 de julho de 2026 (Digital Omnibus on AI). Jornal Oficial, 24 jul. 2026, em vigor desde 27 jul. 2026. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32026R1744

8. DeepSeek. "DeepSeek-V4 Preview: Entering the Era of Affordable Million-Token Context." 24 abr. 2026. https://deepseek.com/en/news/v4-preview/

9. Z.ai. "GLM-5.2: Built for Long-Horizon Tasks." Jun. 2026 (publicado no Hugging Face Blog em 17 jun. 2026). https://z.ai/blog/glm-5.2 e https://huggingface.co/blog/zai-org/glm-52-blog

10. OpenAI. "GPT-6 Astra: A new generation of intelligence." Set. 2026 (atualizações de 22 e 29 set. 2026). https://openai.com/index/gpt-6-astra/; OpenAI API, "GPT-6.1 Sol Model." https://developers.openai.com/api/docs/models/gpt-6.1-sol; The Next Web,

"OpenAI releases GPT-6.1 Sol at a fifth of GPT-6 Astra's token prices." 29 set. 2026. https://thenextweb.com/news/openai-gpt-6-1-sol-price-astra-devday

11. Meta. "Introducing Muse Spark 1.3." Meta AI Research, 2 set. 2026. https://research.meta.ai/blog/introducing-muse-spark-1-3

12. DeepSeek. "Introducing DeepSeek-V4.1-Flash: smarter, faster, more efficient." 10 set. 2026. https://www.deepseek.com/en/news/deepseek-v4-1-flash/

13. SpaceXAI. "Introducing Grok 4.7." https://x.ai/news/grok-4-7; data de lançamento (21 set. 2026) conforme Vercel, "Grok 4.7 now available on AI Gateway." 21 set. 2026. https://vercel.com/changelog/grok-4-7-now-available-and-40-off-on-ai-gateway-fx-eve

14. Anthropic. "Introducing Claude Fable 5.1 and Claude Mythos 5.1." Set. 2026. https://www.anthropic.com/claude-fable-and-mythos-5-1; "Introducing Claude Opus 5.5." 22 set. 2026. https://www.anthropic.com/claude-opus-5-5; "Introducing Claude Sonnet 5.5." 28 set. 2026. https://www.anthropic.com/claude-sonnet-5-5

15. Kavukcuoglu, Koray. "Gemini 4 Argon: our next era of frontier intelligence." Google, 30 set. 2026. https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-4-argon/

16. Artificial Analysis. "Announcing Artificial Analysis Intelligence Index v4.2." 4 set. 2026. https://artificialanalysis.ai/articles/artificial-analysis-intelligence-index-v4-2; "Announcing the Artificial Analysis Intelligence Index v4.3." 7 set. 2026. https://artificialanalysis.ai/articles/artificial-analysis-intelligence-index-v4-3

17. Stanford HAI. "The 2025 AI Index Report." Abr. 2025. Dados sobre queda de custo de inferência (de US$20 para US$0,07 por milhão de tokens, nov. 2022–out. 2024). https://hai.stanford.edu/ai-index/2025-ai-index-report

18. Upwork Research Institute (Kelly Monahan e Gabby Burlacu). "From Burnout to Balance: AI-Enhanced Work Models." 23 jul. 2024. Pesquisa realizada pela Walr para Upwork e Workplace Intelligence, de 16 abr. a 5 maio 2024, com 2.500 respondentes (1.250 executivos C-level, 625

funcionários em tempo integral e 625 freelancers) nos EUA, Reino Unido, Austrália e Canadá. https://www.upwork.com/research/ai-enhanced-work-models

CAPÍTULO 3

Depois do Google

Tese do capítulo: a mudança mais profunda não é trocar um site por outro; é passar de procurar documentos para delegar a construção da resposta.

Terminei o capítulo anterior dizendo que a competência central desta era é aprender a reaprender sem entrar em pânico. Pois bem: uma das primeiras coisas que precisamos reaprender é algo que parecia resolvido havia mais de vinte anos. Como procurar uma resposta.

Eu percebi isso observando gente jovem. Quando uma criança ou um adolescente quer saber alguma coisa hoje, muitas vezes não abre o Google. Abre uma conversa com o ChatGPT, ou com outra inteligência artificial, e pergunta direto. Do jeito que perguntaria a uma pessoa. Sem palavras-chave, sem lista de links, sem escolher em qual página clicar.

Para mim, aquilo foi um pequeno choque. Não porque fosse errado, mas porque era natural demais. Eles não estavam "usando uma tecnologia nova". Estavam simplesmente perguntando. A tecnologia nova era eu tentando entender o que eles faziam.

Não é só impressão. Em pesquisa do Pew Research Center publicada em fevereiro de 2026, 57% dos adolescentes americanos disseram já ter usado chatbots para procurar informação, e 54% para trabalhos escolares.[1] Para uma parte enorme de uma geração, a primeira porta para o conhecimento já não é uma página de resultados. É uma conversa.

Este capítulo é sobre o que acontece quando essa porta muda. E por que a mudança é muito mais profunda do que trocar um site por outro.

1. De palavras-chave, links e páginas para diálogo e síntese

Quem tem a minha idade lembra de pelo menos três maneiras diferentes de procurar informação.

A primeira era física. Em muitas casas, a enciclopédia ficava na estante, volume por volume, em ordem alfabética. Para encontrar um telefone, havia a lista telefônica. Para chegar a um endereço, o mapa de papel, que exigia planejar, interpretar e, muitas vezes, parar e perguntar a alguém na rua. Buscar era um trabalho manual, e a informação envelhecia na estante desde o dia em que tinha sido impressa. Comigo era assim. Lá em Mantena, no interior de Minas Gerais, pesquisa para a escola era a Enciclopédia Barsa e a biblioteca municipal. Eu sou da geração da televisão em preto e branco, do disco de vinil, da máquina de escrever, do mimeógrafo e do telefone de disco.

A segunda maneira foi a da busca na internet. Aprendemos a pensar em palavras-chave. Digitávamos poucas palavras, recebíamos uma lista de páginas e decidíamos em qual link confiar. Era uma habilidade de verdade. Quem sabia buscar bem encontrava em minutos o que outros levavam horas para achar. Durante décadas, pesquisar significou exatamente isso: abrir o Google, digitar, comparar e escolher.

A terceira maneira é a que está chegando agora. Em vez de procurar documentos, a pessoa pede síntese. Em vez de navegar, conversa. Em vez de encontrar apenas informação, pede explicação, comparação, recomendação.

Repare na diferença. No modelo da busca, o sistema encontrava páginas e o trabalho de construir a resposta era seu. Você lia, comparava, desconfiava de uma fonte, confiava em outra e, no fim, juntava as peças. No modelo da conversa, o sistema entrega a resposta já montada. Ele lê por você, compara por você e escreve por você.

É por isso que eu digo que a mudança não é trocar um site por outro. É mudar quem faz o trabalho de montar a resposta. Antes, nós procurávamos. Agora, cada vez mais, nós delegamos.

E as próprias empresas de busca estão mudando junto. Em maio de 2024, o Google começou a mostrar nos Estados Unidos as Visões Gerais criadas por IA: um resumo gerado por inteligência artificial no topo da página de resultados.[2] Em agosto do mesmo ano, o recurso começou a chegar ao Brasil, em português.[3] Em outubro de 2024, a OpenAI lançou a busca dentro do ChatGPT, com respostas acompanhadas de links para as fontes.[4] E, em setembro de 2025, o Google passou a oferecer no Brasil o Modo IA, uma aba de pesquisa em forma de conversa que, segundo a própria empresa, divide a pergunta em subtemas e faz várias buscas ao mesmo tempo em nome do usuário.[5] Não por acaso: quando a interface muda, até quem domina a interface anterior precisa se reinventar. Voltaremos a essa pergunta, a de saber se o Google está ficando para trás, no Capítulo 10.

A vantagem dessa nova forma é enorme, e eu não quero diminuí-la. A conversa permite perguntar como você pensa, e não como a máquina indexa. Permite pedir "explique como se eu tivesse quinze anos" ou "agora explique como se eu fosse a gerente financeira da empresa". Permite fazer a pergunta seguinte sem recomeçar do zero. Para quem nunca se sentiu à vontade com a lógica das palavras-chave, isso é libertador.

Mas toda delegação tem um preço. E o preço aparece com mais clareza quando a inteligência deixa de apenas responder e começa a agir.

2. De respostas para agentes que usam ferramentas e executam tarefas

Se a primeira mudança foi de procurar para perguntar, a segunda é de perguntar para pedir que façam.

Pense em uma tarefa comum do trabalho. Você precisa encontrar três fornecedores para um produto, comparar preços e prazos, preparar um e-mail pedindo proposta e marcar uma reunião com o mais promissor. No mundo da busca, isso significava dezenas de abas abertas, anotações, planilhas e uma tarde inteira. No mundo da conversa, a IA pode ajudar a organizar a pesquisa e escrever o e-mail. No mundo dos agentes, a ideia é outra: você define o objetivo e o sistema planeja as etapas, usa ferramentas e tenta executar.

É isso que significa um agente de inteligência artificial: um sistema que recebe um objetivo, planeja etapas, usa ferramentas e tenta executar ações. A diferença, numa frase simples, é esta: um chatbot responde; um agente se parece mais com um assistente que começa a trabalhar.

Na minha vida de empresária, eu sei exatamente o valor disso. Hoje eu tenho equipes de agentes de inteligência artificial que não apenas respondem: executam. Um constrói e publica os meus sites, outro cuida dos meus e-mails, outro organiza o imobiliário, outro pesquisa ações, outro trabalha com os humanoides e outro escreve comigo os meus livros. Quem administra mais de um negócio aprende que o problema raramente é falta de informação. É falta de tempo para transformar informação em ação.

Por isso, no Teste Ana, que você vai conhecer no Apêndice B, uma das tarefas é justamente esta: executar uma tarefa usando ferramentas, e não apenas explicar como ela poderia ser feita. Hoje, a diferença entre explicar e fazer é uma das fronteiras mais importantes da inteligência artificial.

Só que, quando a IA passa a agir, o erro também muda de natureza. Uma resposta errada pode ser descartada. Uma ação errada pode ter consequência: um e-mail enviado para a pessoa errada, uma compra feita no fornecedor errado, uma reunião marcada no dia errado, um dado confidencial colocado onde não deveria estar. Quanto mais a ferramenta faz, mais importante é saber o que ela fez, com quais informações e com qual autorização.

NÃO CONFUNDA

Busca não é síntese: uma encontra páginas, a outra constrói uma resposta. Chatbot não é agente: um conversa, o outro tenta executar. Agente não é AGI: executar tarefas não significa inteligência geral, consciência ou autonomia ilimitada. E resposta fluente não é resposta correta.

Isso não é motivo para fugir dos agentes. É motivo para usá-los como usaríamos um funcionário novo e muito rápido: com instruções claras, limites definidos, supervisão no começo e confiança construída com o tempo. Ninguém entrega a chave do cofre a um estagiário no primeiro dia, por mais brilhante que ele pareça.

E isso nos leva à questão mais delicada deste capítulo.

3. Quando a conveniência aumenta e o contato com a fonte diminui

Quando alguém abre cinco páginas para pesquisar um assunto, ainda enxerga que existem cinco fontes. Vê que uma é um jornal, outra é um blog, outra é o site de uma empresa interessada em vender, outra é uma universidade. Mesmo sem perceber, compara tons, datas e intenções. A diversidade das fontes fica visível.

Quando recebe uma resposta perfeita em segundos, pode esquecer que aquela frase foi construída por um sistema que interpreta, seleciona e, às vezes, inventa.

Essa é a troca no centro da era depois do Google: a conveniência aumenta e o contato com a fonte diminui.

O nome técnico para quando a IA inventa é alucinação: uma resposta falsa ou não sustentada, apresentada de forma convincente. Eu gosto de explicar assim para quem tem mais de cinquenta anos: é como um funcionário que prefere improvisar uma resposta a admitir que não sabe. O problema não é apenas o erro. É que o erro vem bem escrito, com segurança e no mesmo tom das respostas corretas. Texto elegante não é evidência.

As empresas sabem disso e criaram mecanismos para reduzir o problema. Um dos mais importantes se chama grounding: ligar a resposta a fontes, documentos ou dados verificáveis, em vez de depender apenas da memória geral do modelo. Muitas ferramentas também mostram citações, aqueles pequenos links ao lado das frases. Isso ajuda muito. Reduz erros e permite conferir.

Mas aqui está uma armadilha sutil: ter uma citação não garante que ela foi interpretada corretamente. A fonte pode existir e dizer outra coisa. Pode estar desatualizada. Pode ser uma fonte fraca apresentada com a mesma aparência de uma fonte sólida. A

citação é um convite para verificar, não uma prova de que alguém já verificou.

Os testes independentes mostram por que isso importa. Em março de 2025, o Tow Center for Digital Journalism, da Universidade Columbia, testou oito ferramentas de busca com IA: entregou a elas trechos de reportagens e pediu que identificassem o título, o veículo, a data e o link originais. Somadas, as ferramentas responderam errado a mais de 60% das 1.600 consultas.[6] Em outubro de 2025, um estudo coordenado pela União Europeia de Radiodifusão (EBU) e liderado pela BBC, com 22 emissoras públicas de 18 países, avaliou respostas do ChatGPT, do Copilot, do Gemini e do Perplexity a perguntas sobre notícias. Resultado: 45% tinham pelo menos um problema significativo, e a falha mais comum era justamente a das fontes, com problemas sérios em 31% das respostas.[7] São fotografias datadas, e as ferramentas podem ter melhorado desde então. Mas a lição continua valendo: a citação abre a porta; quem precisa atravessá-la é você.

Existe ainda um segundo custo, menos visível: o que acontece com o nosso próprio raciocínio. Os especialistas chamam de terceirização cognitiva o ato de transferir para uma ferramenta parte do esforço de lembrar, calcular, escrever, navegar ou decidir. Fazemos isso há muito tempo: com a calculadora, com a agenda, com o GPS. Não há nada de errado nisso. O risco aparece quando a pessoa perde a capacidade de compreender e conferir aquilo que delegou.

Na educação, esse risco já foi observado. A OECD concluiu, em 2026, que ferramentas generativas podem elevar a qualidade de uma tarefa sem produzir ganho duradouro de aprendizagem.[8] O trabalho fica melhor enquanto a ferramenta está ali, mas o conhecimento não necessariamente fica na pessoa. Para um

adolescente que faz o trabalho de escola com IA, a pergunta não é só "a resposta está certa?". É também "o que eu aprendi?".

E existe um terceiro custo, que é coletivo. Se cada vez menos pessoas visitam as fontes originais, quem vai sustentar o jornalismo, a pesquisa e os sites que produzem a informação que as próprias inteligências artificiais usam? Os primeiros números já preocupam. Em uma análise do Pew Research Center com dados de navegação de 900 adultos americanos em março de 2025, quem via um resumo de IA no Google clicava em um resultado tradicional em 8% das visitas; quem não via, em 15%. Nos links do próprio resumo, o clique acontecia em apenas 1% das visitas.[9] E, segundo o relatório de tendências de 2026 do Reuters Institute, da Universidade de Oxford, o tráfego vindo da busca do Google para mais de 2.500 sites de notícias caiu 33% entre novembro de 2024 e novembro de 2025, de acordo com dados da Chartbeat. Nem toda essa queda pode ser atribuída à IA, mas os executivos de mídia ouvidos pelo relatório esperam que o tráfego vindo de buscadores caia mais 43% nos próximos três anos.[10] Essa pergunta ainda não tem resposta. Mas ela vai moldar o tipo de internet que teremos daqui a alguns anos.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: a IA é como um colega que leu tudo antes da prova e te conta o resumo. Ajuda muito. Mas, se ele entendeu errado, você vai errar junto, e com a mesma confiança dele. Peça sempre: "de onde você tirou isso?".

Para quem tem 65 anos: o Google era como uma bibliotecária que indicava as prateleiras certas; você ainda lia os livros. A IA é como um assistente que lê os livros e entrega um resumo pronto. É mais rápido, mas você precisa saber quais livros ele leu, e se leu direito.

Então, o que fazemos? Eu não acredito em voltar para a enciclopédia na estante. E não acredito em proibir os jovens de perguntar à IA. Acredito em um hábito simples, que qualquer pessoa pode adotar a partir de hoje:

1. **Peça fontes.** Para qualquer informação que importa, pergunte de onde ela veio.
2. **Abra as fontes que importam.** Não todas; as que vão sustentar uma decisão.
3. **Pergunte o que não foi confirmado.** Uma boa ferramenta deve conseguir dizer onde está a incerteza. Se ela nunca tem dúvida, desconfie.
4. **Calibre o esforço pelo risco.** Uma receita de bolo não exige o mesmo cuidado que um contrato, um diagnóstico ou uma decisão de investimento.

Não é um método sofisticado. É o mesmo bom senso que sempre usamos com pessoas: confiamos mais em quem mostra de onde tirou a informação, e verificamos mais quando o assunto é sério. A diferença é que, agora, precisamos aplicar esse bom senso a uma máquina que fala com muita confiança.

O QUE AINDA NÃO SABEMOS

Como a busca tradicional e a conversa com IA vão conviver nos próximos anos. Se as ferramentas conseguirão reduzir as alucinações a um nível seguro para decisões importantes. Como serão remuneradas as fontes que produzem a informação. E o que acontecerá com a capacidade de pesquisa de uma geração que começou pela resposta pronta.

A resposta pronta é uma das maiores conveniências que a tecnologia já nos deu. Eu a uso em decisões, pesquisa, comunicação e criação, e não pretendo parar. Mas, cada vez que

recebo uma, lembro que delegar a construção da resposta não é o mesmo que delegar o julgamento sobre ela.

E existe um risco que aparece justamente quando nos encantamos com tudo isso: o de tentar usar todas as ferramentas, acompanhar todas as novidades e terminar sem ter mudado nada de verdade. É sobre esse risco que vamos falar no próximo capítulo.

O que mudou / O que significa / O que fazer agora

O que mudou: passamos de procurar documentos para pedir respostas prontas, e estamos começando a passar de pedir respostas para pedir que agentes executem tarefas. Para uma parte crescente dos jovens, a primeira porta para o conhecimento já é uma conversa com a IA.

O que significa: quem constrói a resposta mudou. Isso economiza um tempo extraordinário, mas reduz o contato com as fontes, abre espaço para erros convincentes e pode enfraquecer o aprendizado quando o esforço é inteiramente terceirizado.

O que fazer agora: use a conversa com a IA, mas peça fontes, abra as que importam, pergunte o que não foi confirmado e aumente o cuidado conforme o risco da decisão. A resposta pronta economiza tempo, mas o julgamento sobre a resposta continua sendo seu.

AÇÃO EM DEZ MINUTOS

Escolha uma pergunta real que você faria hoje no Google. Faça a mesma pergunta a uma IA e peça que ela indique as fontes. Abra pelo menos duas delas e confira se dizem mesmo o que a resposta afirma. Anote no seu Diário de Fronteira o que você encontrou: acertos, exageros e erros.

Notas

1. Pew Research Center. "How Teens Use and View AI." 24 fev. 2026. https://www.pewresearch.org/internet/2026/02/24/how-teens-use-and-view-ai/

2. Google. "Google I/O 2024: New generative AI experiences in Search." The Keyword, 14 maio 2024. https://blog.google/products-and-platforms/products/search/generative-ai-google-search-may-2024/

3. Google Brasil (Hema Budaraju). "Novas formas de se conectar com a web com as Visões Gerais criadas por IA." 15 ago. 2024. https://blog.google/intl/pt-br/produtos/explore-e-encontre-respostas/novas-formas-de-se-conectar-com-a-web-com-as-visoes-gerais-criadas-por-ia/

4. OpenAI. "Introducing ChatGPT search." 31 out. 2024. https://openai.com/index/introducing-chatgpt-search/

5. Google Brasil (Robby Stein). "Busca do Google: agora o Brasil já pode usar o Modo IA em português." 8 set. 2025. https://blog.google/intl/pt-br/produtos/busca-do-google-agora-o-brasil-ja-pode-usar-o-modo-ia-em-portugues/

6. Jaźwińska, Klaudia; Chandrasekar, Aisvarya. "AI Search Has a Citation Problem." Tow Center for Digital Journalism / Columbia Journalism Review, 6 mar. 2025. https://www.cjr.org/tow_center/we-compared-eight-ai-search-engines-theyre-all-bad-at-citing-news.php

7. European Broadcasting Union (EBU) e BBC. "News Integrity in AI Assistants: An International PSM Study." Out. 2025. https://www.ebu.ch/news/2025/10/ai-s-systemic-distortion-of-news-is-consistent-across-languages-and-territories-international-study-by-public-service-broadcaste

8. OECD. "OECD Digital Education Outlook 2026: Exploring Effective Uses of Generative AI in Education." 19 jan. 2026. https://www.oecd.org/en/publications/oecd-digital-education-outlook-2026_062a7394-en.html

9. Pew Research Center. "Google users are less likely to click on links when an AI summary appears in the results." 22 jul. 2025. https://www.pewresearch.org/short-reads/2025/07/22/google-users-are-less-likely-to-click-on-links-when-an-ai-summary-appears-in-the-results/

10. Newman, Nic. "Journalism, Media, and Technology Trends and Predictions 2026." Reuters Institute for the Study of Journalism, Universidade de Oxford, jan. 2026 (pesquisa com 280 líderes de mídia de 51 países; dados de tráfego da Chartbeat). https://reutersinstitute.politics.ox.ac.uk/journalism-media-and-technology-trends-and-predictions-2026

CAPÍTULO 4

Velocidade sem direção

Tese do capítulo: quem tenta acompanhar tudo pode terminar sabendo nomes de ferramentas e não tendo transformado nenhum trabalho real.

No fim do capítulo anterior, eu falei de um risco que aparece justamente quando nos encantamos com a inteligência artificial: tentar usar todas as ferramentas, acompanhar todas as novidades e terminar sem ter mudado nada de verdade.

Eu conheço esse risco por dentro. Quem administra mais de uma empresa ao mesmo tempo aprende, às vezes do jeito mais caro, que informação sem prioridade não vira conhecimento. Vira ruído operacional. Relatórios que ninguém lê, reuniões que não decidem, sistemas novos que convivem com os antigos sem substituí-los, e uma equipe que passa mais tempo se adaptando do que produzindo. Eu aprendi o valor da paciência numa mesa de pôquer. No Big 50 da World Series of Poker de 2019, com 28.371 inscritos, passei o Dia 2 recebendo cartas ruins, mão após mão. Os homens ao meu lado repararam. Um deles levantou a sobrancelha e perguntou se eu pretendia desistir para sempre. Eu sorri, desisti de novo e protegi as minhas fichas. Entrei na zona de premiação, e, pela primeira vez, o meu nome apareceu nos resultados oficiais da WSOP. A paciência pode parecer ridícula antes de parecer sábia.

Com a inteligência artificial, esse ruído pode se multiplicar. Há sempre mais uma ferramenta para testar, mais um vídeo para

assistir, mais uma lista de "aplicativos que você precisa conhecer". E é muito fácil confundir movimento com progresso.

Este capítulo fecha a primeira parte do livro com uma proposta simples: antes de acelerar, escolha a direção.

1. A indústria da ansiedade de produtividade

Toda revolução tecnológica cria, ao lado dela, uma pequena indústria de urgência. Com a inteligência artificial, essa indústria ganhou proporções enormes.

Você provavelmente já viu os títulos. "Cinquenta ferramentas de IA que vão mudar sua vida." "Se você não usa isso, já ficou para trás." "Como eu substituí minha equipe inteira em um fim de semana." "O prompt secreto que ninguém está te contando." Cada um deles promete um atalho. Juntos, produzem o contrário: a sensação de que existe um mundo de possibilidades acontecendo sem você, e que a única saída é consumir mais conteúdo, assinar mais serviços e testar mais ferramentas.

Eu chamo isso de indústria da ansiedade de produtividade. Ela não vende produtividade. Vende a sensação de estar correndo atrás dela.

E ela funciona porque toca em sentimentos reais. Vimos no Capítulo 2 que a velocidade da IA mexe com a nossa percepção de tempo, competência e segurança. A indústria da ansiedade explora exatamente essas três feridas. Ela diz que o tempo está acabando, que a sua competência já não basta e que a sua segurança depende de comprar a próxima solução.

O resultado é um tipo muito específico de cansaço. A pessoa conhece o nome de quinze ferramentas, criou contas em metade delas, assistiu a horas de tutoriais, tem três assinaturas que

esqueceu de cancelar... e continua fazendo o trabalho do mesmo jeito que fazia antes. Sabe falar sobre inteligência artificial. Não mudou nada com ela.

Nas empresas, o fenômeno tem outra cara, mas é o mesmo. Um departamento adota uma ferramenta, outro departamento adota outra, um diretor volta de um evento entusiasmado com uma terceira. Ninguém define qual problema está sendo resolvido, ninguém mede o antes e o depois, e ninguém decide quais dados podem ou não ser colocados em cada sistema. Depois de alguns meses, a conclusão é que "a IA não funcionou". Na maioria das vezes, o que não funcionou foi a falta de direção.

Os números mostram essa distância. Na pesquisa global da McKinsey sobre o estado da IA em 2026, com 1.719 participantes, oito em cada dez pessoas que usam IA no trabalho disseram que ela melhorou sua produtividade individual. Mas só 37% atribuíram à IA algum impacto no lucro operacional da organização, praticamente a mesma proporção do ano anterior. A própria McKinsey observa que o pequeno grupo de empresas com resultados expressivos tende a redesenhar os processos de trabalho, em vez de apenas encaixar a IA nos processos que já existiam.[1] É uma consultoria que também vende serviços de transformação, e o dado deve ser lido com isso em mente. Ainda assim, a mensagem combina com tudo o que vivi: ferramenta sem processo não vira resultado.

NÃO CONFUNDA

Conhecer ferramentas não é transformar trabalho. Testar não é adotar. Ter uma assinatura não é ter um resultado. E estar ocupado aprendendo sobre IA não é o mesmo que estar usando IA para criar valor.

Não estou dizendo que experimentar é ruim. Curiosidade é uma das melhores qualidades que alguém pode ter nesta era, e eu sou a primeira a experimentar. O problema é a experimentação sem pergunta. Experimentar para descobrir é inteligente. Experimentar para aliviar a ansiedade é uma esteira: você corre muito e não sai do lugar.

2. Definir prioridades antes de escolher ferramentas

Tudo o que vivi como executiva e empresária me ensinou que boas decisões de tecnologia costumam começar por um problema, e não pela tecnologia. A entrevista com a oficial de IA da Freedom to Belong é um bom exemplo: ela responde a um problema concreto, o do imigrante que precisa treinar, em inglês e em voz alta, as perguntas da entrevista de cidadania do USCIS.

Com a inteligência artificial, vale a mesma regra. A pergunta certa não é "qual é a melhor IA?". É "qual trabalho, na minha vida ou na minha empresa, mais precisa melhorar?".

Comece pelo que consome mais tempo mental. Não necessariamente o que consome mais horas, mas o que mais pesa: o relatório que você adia toda semana, os e-mails difíceis, as propostas repetitivas, a pesquisa que nunca termina, a planilha que precisa ser refeita todo mês. Escreva três dessas tarefas. Esse é, aliás, o primeiro dia do Plano de 30 Dias que está no Apêndice A.

Depois, para cada tarefa, faça algumas perguntas antes de pensar em qualquer ferramenta:

1. **Qual é o objetivo?** O que exatamente precisa ficar melhor: tempo, qualidade, custo, menos erros?

2. **O que entra e o que sai?** Quais materiais e informações a tarefa usa, e qual é o resultado esperado?
3. **Onde está a decisão humana?** Em que ponto alguém precisa julgar, aprovar ou assumir responsabilidade?
4. **Quais dados estão envolvidos?** Existe informação pessoal, confidencial, médica, contratual ou estratégica?
5. **Como vou medir?** Como saberei, daqui a um mês, se ficou melhor?

Essas perguntas parecem burocráticas. Não são. São o que separa a adoção que transforma da adoção que só faz barulho.

Duas delas merecem atenção especial.

A primeira é a dos **dados**. Toda vez que você coloca uma informação em uma ferramenta de IA, precisa saber para onde ela vai e sob quais regras. Não envie segredos comerciais, dados pessoais, informações médicas, contratos ou material controlado sem uma política clara e um ambiente apropriado. Para uma pessoa, isso pode começar com algo simples: uma lista do que nunca deve ser enviado. Para uma empresa, exige regras, ferramentas aprovadas e treinamento. Quanto mais ferramentas você usa, mais difícil fica controlar onde estão os seus dados. Esse é um dos custos invisíveis da dispersão.

A segunda é o **custo de troca**. Cada ferramenta nova cobra um preço que não aparece na assinatura: o tempo para aprender, os hábitos que precisam ser refeitos, os arquivos e históricos que ficam espalhados, a equipe que precisa ser treinada de novo, a configuração que se perde. Trocar de ferramenta toda semana parece agilidade. Muitas vezes, é só recomeço permanente.

Isso não significa nunca trocar. Os modelos estão melhorando rápido, e às vezes a troca vale muito a pena. Significa trocar por um motivo, e não por uma manchete.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: baixar todos os aplicativos de treino da loja não deixa ninguém em forma. Escolher um, treinar de verdade e acompanhar o resultado, sim.

Para quem tem 65 anos: uma cozinha com trinta aparelhos novos não faz o almoço sair melhor. Uma boa faca, uma panela confiável e uma receita bem praticada fazem.

3. A Regra 1-1-1-30

Depois de ver tanta gente, inclusive eu mesma em alguns momentos, perdida no excesso de opções, cheguei a uma regra que gosto de propor a qualquer pessoa ou empresa que esteja começando, ou recomeçando, com a inteligência artificial.

Eu a chamo de Regra 1-1-1-30:

1. **Uma IA principal.** Escolha uma inteligência artificial generalista e aprenda a usá-la bem. Não superficialmente: aprenda os controles, os modos, os limites, o que ela faz bem e onde costuma errar. Uma ferramenta que você conhece profundamente vale mais do que dez que você conhece de nome.
2. **Uma IA especializada.** Só se ela resolver uma limitação real da sua IA principal. Uma ferramenta específica para imagem, para código, para áudio, para planilhas ou para o seu setor. Uma, e apenas quando houver um motivo claro.

3. **Um processo real.** Escolha um único fluxo de trabalho que importa de verdade: pesquisa, e-mail, vendas, estudo, atendimento, orçamento, conteúdo. Use a IA nesse processo, todos os dias, durante um mês. Anote o que melhorou e o que ainda exige você.
4. **Revisão a cada 30 dias.** Uma vez por mês, pare e pergunte: o que ganhei em tempo, custo, qualidade e risco? E decida, para cada ferramenta e para cada processo: manter, abandonar ou aprofundar.

Pode parecer pouco. É de propósito. A regra não existe para limitar a sua curiosidade; existe para dar a ela um lugar onde produzir resultado.

Imagine uma distribuidora industrial com 28 funcionários que gasta tempo demais preparando propostas comerciais. Em vez de comprar cinco ferramentas, ela escolhe uma IA principal, mapeia o processo de propostas, define o objetivo, por exemplo reduzir em 20% o tempo de preparação, decide que desenhos confidenciais não serão enviados a serviços não aprovados e estabelece quem revisa cada proposta antes do envio. Depois de 30 dias, mede o resultado. Se funcionou, aprofunda. Se não funcionou, entende por quê. Em qualquer dos casos, aprendeu algo real. Esse exemplo, aliás, é o mesmo que uso no Laboratório de Prompts do Apêndice C, para mostrar como transformar um pedido vago em uma instrução executável.

Agora imagine uma pessoa sozinha. Uma professora, um advogado, uma corretora de imóveis, um pequeno comerciante. A lógica é idêntica. Uma ferramenta principal, um processo que pesa na semana, uma regra clara sobre o que não compartilhar, e uma revisão no fim do mês. Eu já vivo uma versão dessa regra fora da IA. Administro o meu estudo de pôquer do jeito que administro uma empresa: com agenda, pauta e análise depois. Três vezes por

semana, dou ao jogo trinta minutos de atenção real. Depois de cada sessão, ganhando ou perdendo, reviso três mãos como reviso um negócio que fechou ou desandou: o que eu sabia, quando eu soube, o que faria diferente. Na American Photonics, nunca encerramos um projeto sem essa análise final, porque a lição já está paga.

O mais interessante da regra é o que acontece no segundo e no terceiro mês. Quando você domina uma ferramenta e melhora um processo de verdade, começa a enxergar o próximo processo com outros olhos. A curiosidade deixa de ser ansiosa e passa a ser estratégica. Você não testa uma nova IA porque viu um vídeo; testa porque sabe exatamente qual limitação quer resolver.

E há um efeito colateral precioso: a calma. Quem tem um método não precisa acompanhar tudo. Pode ler as novidades com interesse, sem a sensação de que cada lançamento é uma ameaça pessoal. Lembra do Diário de Fronteira que propus no Capítulo 2? A revisão de 30 dias é o momento perfeito para abri-lo e perguntar se alguma novidade do mês muda alguma coisa no seu trabalho real. Na maioria dos meses, a resposta será "ainda não". E tudo bem.

O QUE AINDA NÃO SABEMOS

Quais ferramentas de hoje ainda existirão em poucos anos. Quanto as IAs generalistas vão absorver das funções que hoje exigem ferramentas especializadas. E como as empresas vão medir, de forma confiável, o ganho real de produtividade trazido pela IA.

Velocidade sem direção é só agitação. Direção sem velocidade, nesta era, também não basta. O que estamos procurando é o encontro das duas: um caminho claro, percorrido com o ritmo que a nova tecnologia permite.

O que mudou / O que significa / O que fazer agora

O que mudou: a quantidade de ferramentas, novidades e promessas de IA cresceu a ponto de criar uma indústria inteira de urgência, que vende mais a sensação de produtividade do que a produtividade em si.

O que significa: tentar acompanhar tudo leva à dispersão, a custos de troca invisíveis, a riscos com dados e a pouco resultado real. A vantagem não está em conhecer mais ferramentas, mas em transformar um trabalho de verdade.

O que fazer agora: aplique a Regra 1-1-1-30: uma IA principal, uma especializada apenas se necessária, um processo real e uma revisão a cada 30 dias para manter, abandonar ou aprofundar. Não teste quinze IAs. Melhore um trabalho real.

AÇÃO EM DEZ MINUTOS

Escreva as três tarefas que mais consomem seu tempo mental. Escolha uma. Para ela, responda em uma linha cada: qual é o objetivo, quais dados estão envolvidos, onde está a decisão humana e como você vai medir o resultado daqui a 30 dias. Esse é o seu primeiro processo real.

Fim da Parte I

Chegamos ao fim da primeira parte deste livro. Começamos com uma cabeça explodindo diante da escala de uma sociedade artificial. Vimos que o mundo começou a mudar toda semana, e que a sensação de atraso permanente mente. Entendemos que a mudança mais profunda não é trocar o Google por outro site, mas delegar a construção da resposta, sem delegar o julgamento. E,

agora, aprendemos que velocidade sem direção não leva a lugar nenhum.

Acordar era o primeiro passo. O segundo é enxergar com clareza.

Porque, para escolher a sua IA principal, você precisa saber o que exatamente está escolhendo. Quando alguém diz "eu uso o ChatGPT", "eu prefiro o Claude" ou "o Gemini é melhor", está falando de uma empresa, de um modelo, de um aplicativo ou de um agente? Essa confusão é muito mais comum do que parece, até entre profissionais experientes. E ela atrapalha quase todas as conversas sobre inteligência artificial.

Na Parte II, vamos organizar esse mapa. Vamos separar quem constrói, o que é treinado, o que aparece na tela e o que realmente age. Vamos entender o que significam open source, open weight, fechado e frontier. E vamos aprender a comparar inteligências sem ser enganados.

Antes de perguntar qual IA é melhor, precisamos descobrir o que é o quê.

Notas

1. McKinsey & Company (Dan Tinkoff, Lieven Van der Veken, Michael Chui e Tara Balakrishnan). "The state of AI in 2026: On the road to ROI." McKinsey Global Survey, ago. 2026 (1.719 participantes). https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai — cobertura independente: The Register, "McKinsey says enterprise AI is finally 'on the road to ROI'", 25 ago. 2026.

PARTE II

O que é o quê

CAPÍTULO 5

Empresa, laboratório, modelo, produto e agente

Tese do capítulo: grande parte da confusão pública desaparece quando o leitor aprende a separar quem constrói, o que é treinado, o que aparece na tela e o que realmente age.

Terminei a primeira parte deste livro dizendo que, antes de perguntar qual IA é melhor, precisamos descobrir o que é o quê. Vou começar confessando por que essa frase me parece tão urgente.

Nas conversas sobre inteligência artificial, eu comecei a prestar atenção nas palavras. E percebi que até profissionais experientes, gente que decide orçamento, contrata equipe e assina contrato, usam "ChatGPT", "GPT" e "OpenAI" como se fossem a mesma coisa. "Eu uso o GPT." "A OpenAI me respondeu errado." "O ChatGPT novo é melhor que o Gemini." Cada uma dessas frases mistura coisas diferentes.

Isso não é preciosismo. No varejo de computadores, eu aprendi cedo que uma coisa é a marca na caixa, outra é o processador que vai dentro, outra é o programa que a pessoa usa na tela. Computadores de marcas diferentes podiam ter o mesmo processador. E o mesmo programa rodava em máquinas de fabricantes diferentes. Quem não entendia essas camadas comparava errado, pagava errado e reclamava da coisa errada.

Com a inteligência artificial, acontece exatamente o mesmo. Só que as camadas são cinco, e os nomes se parecem muito mais entre si.

Este capítulo é um mapa simples dessas cinco camadas: a **empresa**, que é dona e responde pelo negócio; o **laboratório**, que pesquisa e treina; o **modelo**, que é o sistema treinado; o **produto**, que é o que aparece na sua tela; e o **agente**, que é o sistema que recebe um objetivo e tenta executá-lo.

1. OpenAI, Anthropic, Google DeepMind e outros laboratórios

Comecemos por quem constrói.

A OpenAI foi apresentada em dezembro de 2015 como uma organização de pesquisa em inteligência artificial sem fins lucrativos.[1] A estrutura mudou bastante desde então. Em outubro de 2025, a empresa concluiu uma reorganização: a parte sem fins lucrativos passou a se chamar OpenAI Foundation e continua controlando a OpenAI Group, uma empresa de benefício público, ou PBC, na sigla em inglês.[2] No caso da OpenAI, empresa e laboratório quase se confundem: é a mesma organização que pesquisa, treina os modelos e vende os produtos.

A Anthropic foi fundada em 2021 por ex-integrantes da OpenAI, entre eles os irmãos Dario e Daniela Amodei, e também se organiza como empresa de benefício público. O propósito registrado é o desenvolvimento e a manutenção responsáveis de IA avançada para o benefício de longo prazo da humanidade.[3] Aqui, mais uma vez, empresa e laboratório são praticamente a mesma coisa.

No Google, a situação é diferente. O Google é a empresa, que pertence à Alphabet. O laboratório se chama Google DeepMind e nasceu em abril de 2023, quando o Google juntou duas equipes de pesquisa que já existiam, o DeepMind e o Brain, sob o comando de Demis Hassabis. Desde 5 de agosto de 2026, quem dirige o laboratório no dia a dia é Koray Kavukcuoglu; Hassabis passou a presidente do conselho do Google DeepMind e cientista-chefe da Alphabet.[4] Quando alguém diz "o Google lançou um modelo", na prática quem treinou foi o Google DeepMind; quem coloca o modelo na Busca, no Android ou no Gmail são as áreas de produto do Google.

E existem muitos outros. A Meta tem seus próprios modelos. A xAI, criadora do Grok, foi adquirida pela SpaceX em fevereiro de 2026 e hoje assina seus produtos como SpaceXAI.[5] Na China, a Alibaba desenvolve a família Qwen, a Moonshot AI desenvolve o Kimi, e há ainda a DeepSeek e a Z.ai, entre outras. Voltaremos a cada um deles nos próximos capítulos.

Por que essa camada importa para você, que só quer usar a ferramenta? Porque a empresa define coisas que nenhum modelo decide sozinho: o que acontece com os seus dados, quanto você paga, quais regras de uso valem, quem responde legalmente quando algo dá errado e qual é a missão que orienta as escolhas. Duas ferramentas com respostas parecidas podem ter políticas de privacidade, contratos e incentivos muito diferentes.

Quando você escolhe uma IA, também está escolhendo com quem vai fazer negócio.

2. GPT, Claude, Gemini, Kimi, Qwen e outras famílias de modelos

Agora, o que é treinado.

Um modelo de inteligência artificial é o sistema matemático treinado para reconhecer padrões e produzir respostas. Gosto de compará-lo ao motor: o aplicativo é o carro que você dirige, mas é o motor que determina boa parte do que o carro consegue fazer.

Os modelos vêm em famílias, e os nomes contam um pouco da história de cada empresa. "GPT" vem de *Generative Pre-trained Transformer*, uma expressão técnica que aparece no trabalho da OpenAI de 2018 sobre pré-treinamento generativo.[6] Em julho de 2026, a OpenAI lançou a família GPT-5.6, com três modelos de tamanhos e preços diferentes: Sol, Terra e Luna.[7] Dois meses depois, já estava na geração seguinte: o GPT-6 Astra, no início de setembro, o GPT-6 Sol e o GPT-6 Luna, em 22 de setembro, e o GPT-6.1 Sol, uma semana depois.[8] A Anthropic chama seus modelos de Claude. Ao Claude Opus 5, de julho,[9] seguiram-se em setembro o Opus 5.5 e o Sonnet 5.5;[10] em 6 de outubro de 2026, o Opus 5.5 era o modelo mais bem colocado no índice de inteligência da Artificial Analysis.[11] A mesma casa mostra como os nomes podem confundir: segundo a Anthropic, o Claude Fable 5.1 e o Claude Mythos 5.1 são o mesmo modelo, com salvaguardas diferentes, e o Mythos fica restrito a programas de acesso confiável.[12] O Google DeepMind tem a família Gemini, que foi do Gemini 3.6 Flash, publicado em julho,[13] ao Gemini 4 Argon, anunciado em 30 de setembro com acesso inicial restrito.[14] Do lado chinês, a Moonshot AI lançou o Kimi K3 em julho e a equipe Qwen, da Alibaba, apresentou o Qwen3.8-Max em agosto.[15,16]

Não precisa decorar nenhum desses nomes. Eles vão mudar, provavelmente antes de você terminar este livro. O que vale guardar são três ideias.

A primeira: **uma família tem vários tamanhos**. Modelos maiores costumam ser mais capazes e mais caros; modelos

menores, mais rápidos e mais baratos. Escolher o maior para tudo é como usar um caminhão para ir à padaria.

A segunda: **o produto pode trocar de modelo sem trocar de nome**. Quando o ChatGPT foi lançado, em novembro de 2022, ele vinha de um modelo da série GPT-3.5.[17] O nome do aplicativo continuou o mesmo; os modelos por trás dele mudaram muitas vezes desde então. Por isso, a frase "o ChatGPT errou" diz pouco. Qual modelo, em qual versão, em qual configuração?

A terceira: **o mesmo modelo pode estar em produtos diferentes, e o mesmo produto pode usar modelos de empresas diferentes**. O exemplo mais claro é o Microsoft 365 Copilot. Ele é um produto da Microsoft, movido por modelos da OpenAI. Em setembro de 2025, a Microsoft anunciou que passaria a oferecer também modelos Claude, da Anthropic, em algumas funções do Copilot.[18] E, em junho de 2026, lançou uma família de modelos próprios, a MAI.[19] É exatamente a história do processador dentro de computadores de marcas diferentes.

NÃO CONFUNDA

OpenAI é a empresa. GPT é a família de modelos. ChatGPT é o produto. No Google, a empresa é o Google, o laboratório é o Google DeepMind, e "Gemini" é ao mesmo tempo o nome da família de modelos e do aplicativo. Na Anthropic, "Claude" também nomeia modelos e aplicativo. E o Copilot, da Microsoft, é um produto que pode usar modelos de mais de uma empresa.

FOTOGRAFIA DA FRONTEIRA — QUEM É QUEM (6 DE OUTUBRO DE 2026)

OpenAI → família GPT-6 (Astra, Sol, Luna) e GPT-6.1 Sol → ChatGPT, Codex, API e agentes. · Anthropic → Claude Opus 5.5, Sonnet 5.5 e Fable 5.1 (o Mythos 5.1, mesmo modelo, fica restrito) → aplicativo Claude, API e agentes. · Google (laboratório Google DeepMind) → Gemini 4 Argon, com acesso inicial restrito, e Gemini 3.6 Flash → aplicativo Gemini, Busca, Android, Gmail, API. · SpaceXAI → Grok 4.7. · Moonshot AI → Kimi K3. · Alibaba → Qwen3.8-Max. · Microsoft → Copilot, com modelos da OpenAI e da Anthropic, e família própria MAI.

Em julho de 2026, o topo era outro: GPT-5.6, Claude Opus 5 e Claude Fable 5. Nomes e versões mudam rápido: confira a versão atualizada no Atlas Vivo da IA.

3. Chatbots, copilotos, APIs e agentes: quatro experiências diferentes

Por fim, o que aparece na tela e o que realmente age. Aqui está a parte que mais muda a sua vida prática, porque o mesmo modelo pode chegar até você de quatro maneiras muito diferentes.

O chatbot é a janela de conversa. Você pergunta, ele responde. Foi a porta de entrada da IA generativa para o público, e para a maioria das pessoas ainda é a forma principal de contato. A pergunta-chave aqui é a qualidade da resposta: ela é correta, clara, útil e honesta sobre o que não sabe?

O copiloto é a IA embutida em uma ferramenta que você já usa: no editor de texto, na planilha, no e-mail, no programa de apresentações, no ambiente de programação. Você não vai até a

IA; ela vem até o seu trabalho. A pergunta-chave muda: a quais documentos e dados ela tem acesso, e quem controla esse acesso?

A API é a porta pela qual um software chama um modelo. É como um balcão de serviço, com regras definidas para pedidos e respostas. Você talvez nunca veja uma API, mas usa produtos construídos sobre elas o tempo todo: o atendimento automático de uma loja, o resumo de um aplicativo, a tradução de uma plataforma. É por meio de APIs que milhares de empresas colocam modelos que não treinaram dentro dos seus próprios produtos. Aqui, as perguntas são de negócio: preço, velocidade, contrato, e o que acontece com os dados enviados. Usar uma API não significa que os dados ficam automaticamente privados.

O agente é o sistema que recebe um objetivo, planeja etapas, usa ferramentas e tenta executar ações. Em julho de 2025, por exemplo, a OpenAI apresentou o ChatGPT agent, que, segundo a empresa, usa um computador virtual para navegar, preencher formulários, analisar informações e criar planilhas ou apresentações.[20] Na programação, os agentes já têm até avaliações próprias: em julho de 2026, a Artificial Analysis comparava, em seu índice de agentes de programação, sistemas como o Codex, da OpenAI, o Claude Code, da Anthropic, e o Grok Build, da xAI, cada um rodando um modelo por dentro.[21] E os lançamentos de setembro mostraram que o modelo chega primeiro aos agentes: o GPT-6.1 Sol estreou no Codex, no ChatGPT Work e na API antes de chegar à conversa comum do ChatGPT,[8] e o Grok 4.7 foi lançado diretamente no Cursor e no Grok Build.[22] Para agentes, as perguntas mais importantes são de supervisão: o que ele pode fazer sem pedir permissão, como sei o que ele fez, e quem responde pelo erro?

Repare que o motor pode ser o mesmo nos quatro casos. O que muda é o carro, a estrada e quem está ao volante. Um modelo

excelente em um chatbot pode ser frustrante dentro de um copiloto mal integrado. Um modelo mediano pode ser perfeito para uma API que só precisa classificar e-mails rapidamente e barato. E um agente pode decepcionar não porque o modelo é fraco, mas porque recebeu permissões erradas ou instruções vagas.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: pense em um videogame. Existe a empresa que publica o jogo, o estúdio que desenvolve, o motor gráfico que vários jogos diferentes podem usar, o jogo que você baixa e o personagem controlado pelo computador que cumpre missões sozinho. Empresa, laboratório, modelo, produto e agente.

Para quem tem 65 anos: pense em um restaurante. Há o dono, a cozinha, a receita, o prato que chega à mesa ou ao delivery, e o garçom de confiança que, além de servir, resolve as coisas por você. Ninguém culpa a receita quando o garçom erra o pedido. Com a IA, vale a mesma lógica.

O QUE AINDA NÃO SABEMOS

Se as camadas vão se fundir, com empresas controlando tudo, do chip ao agente, ou se vão se separar ainda mais, com produtos que trocam de modelo como quem troca de fornecedor. Quais tipos de experiência, chatbot, copiloto, API ou agente, vão concentrar o maior valor econômico. E como a responsabilidade será dividida quando um agente construído por uma empresa, sobre o modelo de outra, cometer um erro dentro de uma terceira.

Desde que comecei a separar essas camadas, minhas próprias comparações ficaram mais honestas. Quando uma ferramenta me decepciona, eu me pergunto: o problema é o modelo, o produto, a

forma como eu pedi, os dados que eu dei, ou as permissões que eu concedi? Quase sempre a resposta é mais interessante do que "essa IA é ruim".

E existe uma sexta pergunta, que fica para o próximo capítulo: quem pode abrir o motor? Alguns modelos só podem ser usados por meio do produto ou da API da empresa. Outros têm seus pesos disponibilizados para qualquer pessoa baixar e adaptar. Essa diferença, entre aberto e fechado, mexe com preço, soberania, segurança e controle. E, como veremos, ela não tem nada a ver com a outra palavra que vive aparecendo nas manchetes: frontier.

O que mudou / O que significa / O que fazer agora

O que mudou: a inteligência artificial deixou de ser "uma coisa" e virou uma cadeia de camadas: empresas, laboratórios, famílias de modelos, produtos e agentes. Os nomes se repetem e se confundem, e o mesmo modelo pode chegar até você por caminhos muito diferentes.

O que significa: muitas discussões sobre "qual IA é melhor" comparam coisas que não são comparáveis. Política de dados e contrato são decisões da empresa; capacidade é do modelo; experiência é do produto; e execução é do agente, com as permissões que você deu.

O que fazer agora: sempre que avaliar uma ferramenta, identifique a empresa, o modelo e a versão, o tipo de produto e se há um agente agindo em seu nome. Antes de perguntar "qual IA é melhor?", descubra qual camada você está comparando.

AÇÃO EM DEZ MINUTOS

Abra a ferramenta de IA que você mais usa e procure, nas configurações ou no seletor de modelos, qual modelo está respondendo. Escreva três linhas no seu Diário de Fronteira: empresa, modelo e versão, e tipo de experiência (chatbot, copiloto, API ou agente). Se você não conseguir descobrir o modelo, anote isso também. Já é uma informação importante.

Notas

1. OpenAI. "Introducing OpenAI." 11 dez. 2015. https://openai.com/index/introducing-openai/

2. OpenAI. "Built to benefit everyone." 28 out. 2025; ver também "Our structure." https://openai.com/index/built-to-benefit-everyone/ e https://openai.com/our-structure/

3. Anthropic. "Company" e "The Long-Term Benefit Trust." Consultados em 6 out. 2026. https://www.anthropic.com/company e https://www.anthropic.com/news/the-long-term-benefit-trust; sobre a fundação e os fundadores, ver Anthropic, "Anthropic raises $124 million Series A", 28 maio 2021, https://www.anthropic.com/news/anthropic-raises-124-million-to-build-more-reliable-general-ai-systems

4. Google DeepMind. "Announcing Google DeepMind." 20 abr. 2023. https://deepmind.google/blog/announcing-google-deepmind/; Pichai, Sundar. "Google DeepMind: Bringing together two world-class AI teams." Google, 20 abr. 2023. https://blog.google/innovation-and-ai/technology/ai/april-ai-update/; sobre a mudança de comando: Reuters, "Google shakes up AI leadership as DeepMind chief shifts role." 5 ago. 2026. https://www.reuters.com/business/google-shakes-up-ai-leadership-deepmind-chief-shifts-role-2026-08-05/; Whitwam, Ryan. "Google's AI shake-up: DeepMind's Hassabis steps aside, senior scientists depart." Ars Technica, 5 ago. 2026. https://arstechnica.com/gadgets/2026/08/googles-ai-shakeup-deepminds-hassabis-steps-aside-senior-scientists-depart/

5. SpaceX. "xAI joins SpaceX to Accelerate Humanity's Future." 2 fev. 2026. https://www.spacex.com/updates#xai-joins-spacex; xAI. "xAI joins SpaceX." https://x.ai/news/xai-joins-spacex

6. Radford, Alec; Narasimhan, Karthik; Salimans, Tim; Sutskever, Ilya. "Improving Language Understanding by Generative Pre-Training." OpenAI, 2018. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

7. OpenAI. "GPT-5.6: Frontier intelligence that scales with your ambition." 9 jul. 2026. https://openai.com/index/gpt-5-6/

8. OpenAI. "GPT-6 Astra: A new generation of intelligence." Set. 2026 (atualizado em 22 e 29 set. 2026). https://openai.com/index/gpt-6-astra/; OpenAI API, "GPT-6.1 Sol Model." https://developers.openai.com/api/docs/models/gpt-6.1-sol; The Next Web, "OpenAI releases GPT-6.1 Sol at a fifth of GPT-6 Astra's token prices." 29 set. 2026. https://thenextweb.com/news/openai-gpt-6-1-sol-price-astra-devday

9. Anthropic. "Introducing Claude Opus 5." 24 jul. 2026. https://www.anthropic.com/news/claude-opus-5

10. Anthropic. "Introducing Claude Opus 5.5." 22 set. 2026. https://www.anthropic.com/claude-opus-5-5; Anthropic, "Introducing Claude Sonnet 5.5." 28 set. 2026. https://www.anthropic.com/claude-sonnet-5-5

11. Artificial Analysis. "LLM Leaderboard" (Intelligence Index v4.3.2). Consultado em 6 out. 2026. https://artificialanalysis.ai/leaderboards/models

12. Anthropic. "Introducing Claude Fable 5.1 and Claude Mythos 5.1." Set. 2026. https://www.anthropic.com/claude-fable-and-mythos-5-1

13. Google DeepMind. "Gemini 3.6 Flash – Model Card." 21 jul. 2026. https://deepmind.google/models/model-cards/gemini-3-6-flash/

14. Kavukcuoglu, Koray. "Gemini 4 Argon: our next era of frontier intelligence." Google, 30 set. 2026. https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-4-argon/

15. Moonshot AI. "Kimi-K3" (README e licença). Hugging Face, 27 jul. 2026. https://huggingface.co/moonshotai/Kimi-K3; lançamento do serviço em 16 jul. 2026 conforme Artificial Analysis, nota 21.

16. Qwen. "Qwen3.8-Max: A New Bar for Coding and Cowork." 2 ago. 2026. https://qwen.ai/blog?id=qwen3.8

17. OpenAI. "Introducing ChatGPT." 30 nov. 2022. https://openai.com/index/chatgpt/

18. Microsoft. "Expanding model choice in Microsoft 365 Copilot." Microsoft 365 Blog, 24 set. 2025. https://www.microsoft.com/en-us/microsoft-365/blog/2025/09/24/expanding-model-choice-in-microsoft-365-copilot/

19. Suleyman, Mustafa. "Building a hill-climbing machine: Launching seven new MAI models." Microsoft AI, 2 jun. 2026. https://microsoft.ai/news/building-a-hillclimbing-machine-launching-seven-new-mai-models/

20. OpenAI. "Introducing ChatGPT agent: bridging research and action." 17 jul. 2025. https://openai.com/index/introducing-chatgpt-agent/

21. Artificial Analysis. "Four frontier launches in eight days: six labs now field a model above 50 on the Artificial Analysis Intelligence Index." 17 jul. 2026. https://artificialanalysis.ai/articles/four-frontier-launches-in-eight-days-six-labs-now-field-a-model-above-50-on-the-artificial-analysis-intelligence-index

22. SpaceXAI. "Introducing Grok 4.7." Set. 2026. https://x.ai/news/grok-4-7; data de lançamento (21 set. 2026) conforme Vercel, "Grok 4.7 now available on AI Gateway." 21 set. 2026. https://vercel.com/changelog/grok-4-7-now-available-and-40-off-on-ai-gateway-fx-eve

CAPÍTULO 6

Open source, open weight, fechado e frontier

Tese do capítulo: acesso e capacidade são eixos diferentes: "open" descreve controle; "frontier" descreve proximidade da fronteira de desempenho.

No capítulo anterior, deixei uma pergunta no ar: quem pode abrir o motor?

Sempre que explico isso a alguém, acabo voltando ao carro. Não porque eu seja especialista em mecânica, mas porque quase todo mundo já viveu as quatro situações que vou descrever. Você pode alugar um carro: dirige, chega aonde precisa, mas não abre o capô, não troca peças e devolve do jeito que a locadora exige. Pode comprar um carro com o motor montado e levá-lo à sua oficina: ajusta, troca componentes, adapta para o seu uso, mas não recebeu o projeto da fábrica. E pode, em um caso mais raro, receber o carro, o motor, o projeto completo, a lista de materiais e a descrição do processo de fabricação, com permissão para estudar, modificar e construir o seu.

Essas são, de forma bem simplificada, as diferenças entre um modelo fechado, um modelo open weight e um sistema open source.

E existe uma quarta situação, que não tem nada a ver com as outras três: o carro de Fórmula 1. Ele representa o limite do desempenho naquele momento. Pode ser de uma equipe que

esconde tudo ou de uma que publica seus projetos. Ser o mais rápido da pista e ser aberto à oficina são perguntas diferentes.

É isso que chamamos de **frontier**: os modelos que estão entre os mais avançados disponíveis em determinado momento. E é por isso que este capítulo trata de dois eixos, e não de um. "Open" responde quem controla. "Frontier" responde até onde a tecnologia consegue chegar.

1. O que realmente significa usar, estudar, modificar e compartilhar

A expressão "open source" tem dono intelectual. Há décadas, a Open Source Initiative, conhecida como OSI, mantém a definição de software de código aberto mais usada no mundo. Em outubro de 2024, depois de um processo de consulta pública que durou cerca de um ano, ela publicou a primeira versão estável de uma definição específica para inteligência artificial: a Open Source AI Definition 1.0.[1,2]

A definição se apoia em quatro liberdades. Um sistema de IA open source precisa permitir que qualquer pessoa:

Use o sistema para qualquer finalidade, sem pedir permissão.

Estude como ele funciona e inspecione seus componentes.

Modifique o sistema para qualquer finalidade, inclusive mudando o que ele produz.

Compartilhe o sistema com outras pessoas, com ou sem modificações, para qualquer finalidade.[1]

Parece simples, mas há uma condição que muda tudo: para exercer essas liberdades, é preciso ter acesso à forma preferida para fazer modificações. No caso da IA, a OSI diz que isso inclui

três coisas. Os **parâmetros**, ou pesos, que são o resultado final do treinamento. O **código** completo usado para treinar e rodar o sistema. E **informação suficientemente detalhada sobre os dados** de treinamento, a ponto de uma pessoa qualificada conseguir construir um sistema substancialmente equivalente.[1]

Repare em um detalhe: a OSI não exige que todos os dados sejam publicados, porque alguns não podem ser compartilhados legalmente. Mas exige que sejam descritos: de onde vieram, como foram selecionados, como foram filtrados e onde é possível obtê-los.[1,3]

Na linguagem do carro: open source é receber o motor, o projeto, a lista de materiais e o manual do processo de fabricação. Não basta poder dirigir. É preciso poder entender e reconstruir.

Quantos sistemas cumprem essa régua? Poucos. Quando testou a definição, a OSI citou como exemplos que passaram em sua validação o Pythia, da EleutherAI, o OLMo, do Allen Institute for AI, o Amber e o CrystalCoder, do LLM360, e o T5, do Google, ressaltando que isso não era uma certificação.[2] Nenhum deles é o nome que aparece nas manchetes desta semana.

2. Por que open weight não é automaticamente open source

Agora, o caso mais comum, e o que mais confunde.

Um modelo **open weight** é aquele cujos parâmetros finais treinados foram disponibilizados para download. Você recebe o "cérebro pronto": pode rodá-lo nos seus próprios servidores, adaptá-lo aos seus dados e ajustá-lo para uma tarefa específica. É o motor montado, entregue na sua oficina.

O que normalmente falta? Segundo a própria OSI, o código de treinamento, o conjunto de dados e a transparência completa sobre a composição desses dados. Sem isso, fica difícil reproduzir, auditar ou entender a fundo como o modelo foi construído.[3] A OSI reconhece que pesos abertos são um avanço em relação a sistemas totalmente fechados, mas insiste: não são a mesma coisa que open source.[3]

Há ainda uma segunda camada: a licença. Baixar não significa poder fazer qualquer coisa. Três exemplos reais mostram como isso varia.

A Meta popularizou os pesos abertos com a família Llama, mas a OSI afirma que a licença do Llama não é open source, porque restringe usos e impõe condições conforme o porte de quem usa.[4]

A Moonshot AI publicou em julho de 2026 os pesos completos do Kimi K3 e o descreve como um modelo open weight. A licença própria permite usar, modificar, distribuir e vender, mas impõe condições comerciais: empresas que oferecem o modelo como serviço e passam de 20 milhões de dólares de receita em doze meses precisam de um acordo separado com a Moonshot.[5,6]

Já a OpenAI, conhecida pelos modelos fechados, lançou em agosto de 2025 dois modelos open weight, o gpt-oss-120b e o gpt-oss-20b, sob a licença Apache 2.0, acompanhados de uma política de uso.[7] E a chinesa Z.ai publicou os pesos do GLM-5.2 em junho de 2026 sob licença MIT, que ela própria chama de licença open source.[8]

Uma licença permissiva sobre os pesos é um passo importante. Mas, pela régua da OSI, ela é necessária e não suficiente: sem código de treinamento e informação detalhada sobre os dados, o sistema continua sendo open weight.

NÃO CONFUNDA

Disponível para download não significa open source. Open weight significa que os pesos foram liberados; open source, na definição da OSI, exige também código e informação sobre os dados, com liberdades de usar, estudar, modificar e compartilhar. E gratuito não significa barato: quem roda um modelo aberto paga pela infraestrutura, pela energia, pela segurança e pelas pessoas que cuidam dele.

E o **modelo fechado**? É aquele controlado pela empresa e acessado por aplicativo ou API, sem distribuição dos pesos. É o carro alugado. Você não vê o motor, mas também não precisa cuidar dele. Muitas empresas preferem isso: menos trabalho técnico, atualizações constantes e um fornecedor responsável. O preço é a dependência: a empresa decide preços, versões, regras e disponibilidade.

A diferença tem ficado mais importante, não menos. O AI Index 2026, da Universidade Stanford, registrou que os modelos mais capazes estão entre os menos transparentes: o índice que mede a transparência das grandes empresas de IA sobre dados, computação, capacidades e riscos caiu de uma média de 58 para 40 pontos.[9]

3. Como um modelo aberto pode ser frontier e um modelo fechado pode ser mediano

Agora vamos cruzar os dois eixos.

Durante muito tempo, havia uma suposição confortável: o que é fechado é melhor, o que é aberto é a versão popular, um degrau abaixo. Os dados mostram uma história mais interessante.

O AI Index 2026 acompanhou essa distância no Arena Leaderboard, uma plataforma de comparação baseada na preferência de usuários. Em maio de 2023, o melhor modelo fechado superava o melhor modelo aberto por 15,2%. Em agosto de 2024, a distância tinha caído para 0,5%. Depois, voltou a crescer com novos modelos fechados de fronteira: em março de 2026, o melhor fechado, o Claude Opus 4.6, liderava o melhor aberto, o GLM-5, por 3,4%.[10] Ou seja: os fechados ainda lideravam naquele recorte, mas por uma margem muito menor do que alguns anos antes.

Poucos meses depois, um retrato independente mostrou outra fotografia, que hoje já é registro histórico. Em 17 de julho de 2026, no índice de inteligência da Artificial Analysis, o Kimi K3, da Moonshot AI, aparecia em terceiro lugar geral, com 57 pontos, à frente do Claude Opus 4.8, um modelo fechado, com 56.[11] Pouco depois, os pesos do Kimi K3 foram publicados.[5,6] No mesmo índice, o GLM-5.2, com pesos abertos, empatava em 51 pontos com dois modelos fechados: o GPT-5.6 Luna, da OpenAI, e o Muse Spark 1.1, da Meta.[11]

Em outubro, a fotografia mudou de novo, e a régua também. A Artificial Analysis lançou novas versões do índice em setembro, por isso as notas de julho e de outubro não se comparam diretamente; o que se compara é a posição. Em 6 de outubro de 2026, o melhor modelo fechado, o Claude Opus 5.5, tinha 58 pontos. O melhor aberto, o MiMo-V2.6-Pro, da Xiaomi, tinha 46, empatado com um fechado, o Grok 4.7. Logo atrás vinham o GLM-5.3, com 45, e o Kimi K3, com 44.[12] Com a onda de lançamentos fechados de setembro, a distância entre o topo fechado e o topo aberto voltou a crescer. Mas o mesmo retrato mostra o outro lado: o GPT-5.6 Terra, fechado e da geração

anterior, aparecia com 42 pontos, abaixo de vários modelos abertos.[12] Fechado não é sinônimo de melhor.

Esse último detalhe merece atenção. A Meta, que ajudou a popularizar os pesos abertos, mantém o Muse Spark como modelo proprietário, sem liberar pesos; a versão 1.3, de setembro de 2026, chegou pela API e pelos produtos da própria Meta.[13] Nem as empresas cabem em uma única caixa: a OpenAI tem modelos fechados e abertos; a Meta também.

Então um modelo aberto pode, sim, estar na fronteira. E um modelo fechado pode ser mediano, ou simplesmente inadequado para o seu caso: o fato de uma empresa guardar os pesos não garante que aquele modelo seja o melhor, o mais seguro ou o mais barato para a sua tarefa.

FOTOGRAFIA DA FRONTEIRA — ABERTO E FECHADO (6 DE OUTUBRO DE 2026)

Artificial Analysis, índice v4.3.2: melhor fechado, Claude Opus 5.5, 58 pontos. Melhores abertos: MiMo-V2.6-Pro (Xiaomi), 46, empatado com o Grok 4.7 (fechado); GLM-5.3, 45; Kimi K3, 44. GPT-5.6 Terra (fechado), 42.

Registro histórico, outra versão do índice (17 jul. 2026): Kimi K3, 57 pontos, em 3.º lugar, à frente do Claude Opus 4.8 (fechado), 56. AI Index 2026: em março de 2026, o melhor fechado liderava o melhor aberto por 3,4% no Arena Leaderboard; em agosto de 2024, por 0,5%.

Notas de versões diferentes do índice não se comparam. Nomes, notas e licenças mudam rápido: confira o Atlas Vivo da IA.

Há também uma dimensão geopolítica. O AI Index 2026 registrou que a distância de desempenho entre os melhores modelos americanos e chineses praticamente desapareceu: em março de 2026, o melhor modelo americano liderava por 2,7%.[9,10] Boa parte

da força chinesa tem chegado ao mundo em forma de pesos abertos, como o DeepSeek V4 Preview, de abril de 2026, sob licença MIT, ou o Tencent Hy4 preview, de agosto de 2026, sob licença Apache 2.0.[14,15] Isso afeta diretamente países como o Brasil. Um governo, um hospital ou uma empresa que pode rodar um modelo aberto nos próprios servidores ganha controle sobre dados e custos. Mas continua dependendo de chips, energia e talento técnico que nem sempre estão em casa.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: fechado é jogar um jogo on-line que você não pode modificar. Open weight é poder instalar o jogo e criar mods, mas sem o código-fonte original. Open source é ter o código, as ferramentas e as instruções para refazer o jogo do zero. Frontier é o jogo com os gráficos mais avançados do ano, e ele pode ser de qualquer um dos três tipos.

Para quem tem 65 anos: pense em uma receita de família. Fechado é comer no restaurante. Open weight é receber o bolo pronto, que você pode recortar, rechear e servir do seu jeito. Open source é receber a receita completa, a lista de ingredientes e o passo a passo. Frontier é o melhor bolo da cidade naquele mês, seja de quem for.

O QUE AINDA NÃO SABEMOS

Se os modelos abertos vão alcançar e manter a liderança absoluta ou continuar alguns passos atrás dos fechados. Se governos vão restringir a publicação de pesos de modelos muito capazes por razões de segurança. Como o direito vai tratar os pesos: a própria OSI reconhece que a natureza jurídica dos parâmetros ainda vai se esclarecer com o tempo.[16] E se as licenças "quase abertas", com condições por porte ou receita, vão virar o padrão do mercado.

Eu gosto de olhar para essa discussão como olhava para os canais de venda no varejo. Nenhum canal era bom ou ruim em si. Dependia do produto, do cliente, da margem e do controle que eu precisava ter. Quando assumi como CEO da American Photonics, em 2014, a empresa dependia de um modelo concentrado num único distribuidor nacional. Reestruturei a distribuição para canais diversificados no mundo inteiro, reposicionei a marca e modernizei o marketing. Passamos a atender clientes nas Américas, na Europa, na Ásia, no Oriente Médio e na Austrália, e, em quatro anos, a nossa participação no mercado global saiu de cerca de 3% para 15%. Foi exatamente esse tipo de escolha: depender de um canal só ou construir vários. Com modelos de IA, é igual. Um escritório pequeno pode preferir o carro alugado. Uma empresa que lida com dados sensíveis pode preferir ter o motor na própria oficina. E um pesquisador pode precisar do projeto completo.

Saber em qual eixo você está é o primeiro passo. O segundo é aprender a medir. Porque, se "frontier" depende de quem chega mais longe, alguém precisa medir a distância, e nem toda medição é honesta ou útil para você. É disso que trata o próximo capítulo.

O que mudou / O que significa / O que fazer agora

O que mudou: "aberto" deixou de significar uma única coisa. Hoje há modelos fechados, modelos open weight com licenças muito diferentes e poucos sistemas que cumprem a definição completa de open source da OSI. Ao mesmo tempo, modelos abertos chegaram perto da fronteira e já passaram à frente de modelos fechados importantes, ainda que o topo absoluto, em outubro de 2026, continuasse fechado.

O que significa: aberto ou fechado descreve quem pode controlar, estudar e operar a tecnologia. Frontier descreve até onde ela consegue chegar. São perguntas diferentes, e misturá-las leva a decisões ruins.

O que fazer agora: antes de escolher um modelo, pergunte separadamente: eu preciso de controle sobre o motor ou só de chegar ao destino? Qual licença se aplica? Quanto custa rodar? Qual é o desempenho para a minha tarefa? Open ou fechado não responde sozinho se o modelo é bom, seguro, barato ou adequado para você.

AÇÃO EM DEZ MINUTOS

Escolha um modelo que você usa ou que apareceu em uma notícia recente. Responda três perguntas no seu Diário de Fronteira: os pesos são públicos? Se sim, qual é a licença e que condições ela impõe? Ele está entre os primeiros em algum ranking independente, e em que data? Se você não encontrar uma das respostas, anote isso também: transparência também é uma informação.

Notas

1. Open Source Initiative. "The Open Source AI Definition – 1.0." 2024. https://opensource.org/ai/open-source-ai-definition

2. Open Source Initiative. "Open Source AI" (página da definição, incluindo a fase de validação) e "OSI's Open Source AI Definition 1.0 Unveiled at All Things Open." 28 out. 2024. https://opensource.org/ai e https://opensource.org/press-mentions/osis-open-source-ai-definition-1-0-unveiled-at-all-things-open; InfoWorld, "OSI unveils Open Source AI Definition 1.0." 28 out. 2024. https://www.infoworld.com/article/3593266/osi-unveils-open-source-ai-definition-1-0.html

3. Open Source Initiative. "Open Weights: not quite what you've been told." Consultado em 6 out. 2026. https://opensource.org/ai/open-weights

4. Open Source Initiative. "Meta's LLaMa license is still not Open Source." https://opensource.org/blog/metas-llama-license-is-still-not-open-source

5. Moonshot AI. "Kimi-K3" (README e "Kimi K3 License"). Hugging Face, publicado em 27 jul. 2026. https://huggingface.co/moonshotai/Kimi-K3 e https://huggingface.co/moonshotai/Kimi-K3/blob/main/LICENSE

6. Unite.AI. "Moonshot Opens Kimi K3 Weights Under a Revenue-Tiered License." 27 jul. 2026. https://www.unite.ai/moonshot-opens-kimi-k3-weights-under-a-revenue-tiered-license/

7. OpenAI. "Introducing gpt-oss." 5 ago. 2025. https://openai.com/index/introducing-gpt-oss/; OpenAI Help Center, "OpenAI open-weight models (gpt-oss)." https://help.openai.com/en/articles/11870455

8. Z.ai. "GLM-5.2: Built for Long-Horizon Tasks." Hugging Face Blog, 17 jun. 2026. https://huggingface.co/blog/zai-org/glm-52-blog; repositório com licença MIT: https://huggingface.co/zai-org/GLM-5.2

9. Stanford HAI. "Inside the AI Index: 12 Takeaways from the 2026 Report." 2026. https://hai.stanford.edu/news/inside-the-ai-index-12-takeaways-from-the-2026-report

10. Stanford HAI. "The 2026 AI Index Report", cap. 2, "Technical Performance." 2026. https://hai.stanford.edu/assets/files/ai_index_report_2026_chapter_2_technical.pdf

11. Artificial Analysis. "Four frontier launches in eight days: six labs now field a model above 50 on the Artificial Analysis Intelligence Index." 17 jul. 2026. https://artificialanalysis.ai/articles/four-frontier-launches-in-eight-days-six-labs-now-field-a-model-above-50-on-the-artificial-analysis-intelligence-index

12. Artificial Analysis. "LLM Leaderboard" (Intelligence Index v4.3.2; classificação de pesos abertos e fechados da própria plataforma). Consultado em 6 out. 2026. https://artificialanalysis.ai/leaderboards/models

13. Meta. "Introducing Muse Spark: Meta's Most Powerful Model Yet." Abr. 2026. https://about.fb.com/news/2026/04/introducing-muse-spark-meta-superintelligence-labs/; Ars Technica, "Meta's Superintelligence Lab unveils its first public model, Muse Spark." Abr. 2026. https://arstechnica.com/ai/2026/04/metas-superintelligence-lab-unveils-its-first-public-model-muse-spark/; Meta, "Introducing Muse Spark 1.3." Meta AI Research, 2 set. 2026. https://research.meta.ai/blog/introducing-muse-spark-1-3

14. DeepSeek. "DeepSeek-V4 Preview: Entering the Era of Affordable Million-Token Context." 24 abr. 2026. https://deepseek.com/en/news/v4-preview/; pesos e licença MIT: https://huggingface.co/deepseek-ai/DeepSeek-V4-Pro

15. Tencent. "Tencent Releases and Open-Sources Tencent Hy4 preview." 28 ago. 2026. https://www.tencent.com/tencent-releases-and-open-sources-tencent-hy4-preview/

16. Open Source Initiative. "OSAID FAQs." https://opensource.org/ai/faq

CAPÍTULO 7

Como comparar inteligências sem ser enganado

Tese do capítulo: um ranking mede um recorte; a escolha real depende de tarefa, dados, preço, velocidade, ferramentas e confiança.

Terminei o capítulo anterior dizendo que, se "frontier" depende de quem chega mais longe, alguém precisa medir a distância. E que nem toda medição é honesta ou útil para você.

Toda semana, alguém me manda um gráfico. Barras coloridas, um modelo novo no topo, uma legenda triunfante: "o mais inteligente do mundo". Na semana seguinte, chega outro gráfico, com outro vencedor. Durante um tempo, eu levava esses gráficos muito a sério. Depois, comecei a fazer o que sempre fiz no varejo diante de uma tabela de vendas bonita demais: perguntar o que exatamente estava sendo medido, por quem, em qual período e com quais exclusões.

A resposta quase sempre muda a história.

Foi assim que nasceu o que, neste livro, chamo de Teste Ana. Em vez de confiar apenas nos rankings, passei a pegar problemas reais do meu trabalho, como pesquisar um tema, ler um documento longo, calcular uma margem, escrever uma comunicação difícil e executar uma tarefa, e entregar exatamente o mesmo problema a diferentes IAs.

Antes de mostrar o teste, porém, precisamos entender o que os rankings medem, e o que eles não medem.

1. Benchmark é prova, não vida real

Um benchmark é um teste usado para comparar modelos em tarefas definidas. Para quem tem menos de 30, é uma prova aplicada a várias IAs. Para quem tem mais de 50, é como comparar carros em pista, consumo e frenagem. É útil. Mas ninguém compra um carro apenas pelo tempo de frenagem medido em uma pista perfeita, com piso seco e piloto profissional.

Os benchmarks têm três limites que todo leitor deveria conhecer.

O primeiro é o recorte. Cada prova mede uma habilidade específica: matemática, programação, raciocínio científico, leitura de documentos, uso de ferramentas. Primeiro lugar em uma prova não significa melhor em tudo.

O segundo é a contaminação. Quando as perguntas de uma prova, ou perguntas muito parecidas, acabam nos dados de treinamento, o modelo pode estar lembrando em vez de raciocinar. Pesquisadores da Scale AI testaram isso de forma engenhosa. Criaram o GSM1k, um conjunto novo de problemas de matemática do nível escolar, desenhado para ter o mesmo estilo e a mesma dificuldade de um benchmark muito conhecido, o GSM8k. Alguns modelos tiveram quedas de precisão de até 8% na prova nova, com sinais de que tinham memorizado parcialmente a antiga. Ao mesmo tempo, muitos modelos de fronteira mostraram poucos sinais desse problema.[1] A lição não é "todo benchmark mente". É: uma nota alta em uma prova velha e famosa vale menos do que parece.

O terceiro é a saturação. Os modelos estão melhorando tão rápido que muitas provas deixam de separar os melhores. O AI Index 2026 relatou, por exemplo, que a taxa de sucesso de agentes em tarefas reais medidas pelo Terminal-Bench subiu de 20% em 2025 para 77,3%.[2] Quando todos se aproximam do teto, a prova perde poder de distinção. A Artificial Analysis chegou a retirar uma avaliação, o IFBench, de seu índice principal justamente porque ela já não distinguia suficientemente os modelos de fronteira. Em setembro de 2026, na versão 4.2, fez o mesmo com o GPQA Diamond, uma prova de ciência de nível de pós-graduação, por considerá-la saturada.[3]

Há também um quarto limite, mais humano: a preferência. O ranking mais citado do setor durante anos foi o Chatbot Arena, hoje LMArena, criado por pesquisadores ligados à Universidade da Califórnia em Berkeley. A ideia é elegante. Uma pessoa escreve uma pergunta, recebe duas respostas de modelos anônimos e vota na que prefere. Com centenas de milhares de votos, a plataforma calcula uma classificação.[4] É uma medida valiosa de como as pessoas reagem às respostas. Mas preferência não é o mesmo que precisão. Uma resposta mais simpática, mais longa ou mais bem formatada pode ganhar o voto sem estar mais correta.

Um benchmark responde se o modelo passou na prova. Só o seu trabalho responde se ele serve para você.

2. A diferença entre alegação da empresa e avaliação independente

Quando uma empresa lança um modelo, ela publica números. Isso é normal e útil. Mas é preciso lembrar quem escolheu as provas,

quem configurou o teste, quais comparações entraram no gráfico e quais ficaram de fora. Neste livro, resultados divulgados pelas próprias empresas são tratados como o que são: alegações da empresa.

Um episódio de abril de 2025 mostra por quê. A Meta lançou os modelos Llama 4 e destacou que uma "versão experimental de chat" do Llama 4 Maverick havia alcançado uma pontuação muito alta no LMArena. Dias depois, o próprio LMArena declarou publicamente que a interpretação da Meta sobre suas regras não correspondia ao que a plataforma esperava, e que a Meta deveria ter deixado mais claro que aquela versão era um modelo customizado para otimizar a preferência humana. A plataforma publicou mais de 2.000 comparações para revisão e atualizou suas regras.[5,6] A Meta, por sua vez, afirmou que experimenta com todo tipo de variante customizada.[6] O ponto, para nós, não é escolher um vilão. É perceber que a versão avaliada pode não ser exatamente a versão que você vai usar.

No mesmo mês, um grupo de pesquisadores publicou um estudo chamado "The Leaderboard Illusion", a ilusão do ranking. Eles relataram que alguns fornecedores puderam testar várias variantes privadas antes de um lançamento e escolher qual divulgar. No caso mais extremo identificado, foram 27 variantes privadas da Meta antes do Llama 4. O estudo também estimou que modelos fechados recebiam mais comparações na plataforma do que os abertos, o que dá mais dados a alguns participantes do que a outros.[7]

E as avaliações independentes? Elas ajudam muito, mas também têm escolhas próprias. A Artificial Analysis, que citei nos capítulos anteriores, afirma conduzir de forma independente todas as avaliações do seu índice. Na versão 4.3 do índice, lançada em 7 de setembro de 2026 e consultada em outubro, já na revisão 4.3.2,

ele combinava dez avaliações em quatro grupos (agentes, programação, conhecimento geral e raciocínio científico), e 45% do peso vinha de provas com perguntas ou respostas privadas, para reduzir a possibilidade de as empresas treinarem para a prova.[8] Isso é bom. Mas significa também que o índice muda de versão, e que uma nota de julho não é diretamente comparável a uma nota calculada com outra versão.

Há ainda os documentos que acompanham os modelos. Um **model card** é uma ficha técnica que descreve um modelo: para que foi feito, como foi avaliado, quais são seus limites. A ideia foi proposta em um artigo de 2019 por pesquisadores que defendiam mais transparência na publicação de modelos de aprendizado de máquina.[9] Um **system card** descreve o sistema completo, com produto, proteções e políticas. A própria OpenAI explicou a diferença ao publicar seus modelos abertos gpt-oss: chamou o documento de model card, e não de system card, porque esses modelos seriam usados dentro de muitos sistemas diferentes, construídos por outras pessoas.[10]

Esses documentos são valiosos e merecem leitura. Mas continuam sendo documentos da empresa.

NÃO CONFUNDA

Benchmark é uma prova. Ranking é uma classificação construída a partir de uma ou mais provas. Alegação da empresa é o resultado que ela escolheu divulgar. Avaliação independente é feita por terceiros, com metodologia própria e também com limites. Nenhum deles, sozinho, responde se o modelo é o melhor para o seu trabalho.

FOTOGRAFIA DA FRONTEIRA — O TERMÔMETRO TAMBÉM MUDA

Julho de 2026 (registro histórico): em 17 de julho, a Artificial Analysis colocava o Claude Fable 5 em primeiro lugar no seu índice de inteligência, com 60 pontos, seguido de perto por GPT-5.6 Sol e pelo Kimi K3, este com 57.[11]

Setembro: o índice ganhou duas versões novas em quatro dias: a 4.2, em 4 de setembro, que retirou o GPQA Diamond, e a 4.3, em 7 de setembro, com 45% do peso em provas privadas.[3,8]

6 de outubro de 2026 (versão 4.3.2): Claude Opus 5.5 em primeiro, com 58 pontos; Kimi K3 com 44.[8] O Kimi não piorou: mudou a régua. Antes de comparar duas notas, confira se vieram da mesma versão do mesmo índice.

Notas, versões e metodologias mudam rápido: confira a versão atualizada no Atlas Vivo da IA.

3. O Teste Ana: oito tarefas iguais, observadas com critérios humanos

Nada disso significa ignorar rankings. Eu os uso como um corretor usa o histórico de preços de um bairro: para saber por onde começar, não para decidir sozinho. A decisão vem do teste no seu próprio terreno.

O Teste Ana tem uma regra simples: as mesmas tarefas são entregues a cada modelo, com o mesmo material e o mesmo prazo. E cada teste registra a versão do modelo, a data, o preço, as ferramentas permitidas e se o resultado veio do aplicativo, da API ou de um agente. Sem esse registro, a comparação vira opinião.

As oito tarefas são estas:

1. Explique o mesmo tema para uma pessoa de 15 anos e para uma pessoa de 70, sem infantilizar nenhuma delas.
2. Pesquise uma informação atual, apresente fontes e identifique claramente o que não pôde ser confirmado.
3. Leia um documento longo e encontre riscos, contradições, números e ações que realmente importam.
4. Escreva uma comunicação empresarial difícil de forma firme, clara e capaz de preservar o relacionamento.
5. Calcule custo, margem, desperdício e cenários; mostre as premissas para que outra pessoa consiga conferir.
6. Crie um plano realista para uma empresa de uma pessoa, separando automação de responsabilidade humana.
7. Execute uma tarefa usando ferramentas, e não apenas explique como ela poderia ser feita.
8. Reconheça quando faltam dados, quando uma fonte é fraca e quando a resposta é apenas uma hipótese.

Repare que nenhuma delas é uma pegadinha. São tarefas da vida real: pesquisa, documentos, margem, comunicação e execução. A oitava é, para mim, a mais importante. Uma IA que sabe dizer "não sei" ou "esta fonte é fraca" vale mais do que uma que responde tudo com a mesma segurança.

Como avaliar? Com critérios humanos, que você pode pontuar de 1 a 5:

Precisão factual e matemática. **Fontes**: existem, são boas e dizem o que a resposta afirma? **Clareza** para o público pedido. **Obediência** às instruções e preservação do contexto. **Utilidade** prática do resultado final. **Execução** com ferramentas sem perder o controle. **Privacidade, segurança** e adequação ao uso empresarial.

Algumas recomendações práticas que aprendi aplicando o teste. Use material real, mas retire dados confidenciais antes, ou use uma versão da ferramenta com proteções contratuais adequadas. Não ajude mais um modelo do que outro: o mesmo pedido, palavra por palavra. Faça o teste mais de uma vez, porque a mesma IA pode responder de forma diferente ao mesmo pedido. E, se possível, peça a outra pessoa que avalie as respostas sem saber qual modelo produziu cada uma, como em uma degustação às cegas.

O vencedor de um teste não recebe o título de "melhor IA". Recebe o título de melhor desempenho naquela tarefa, naquela versão, naquele dia e naquela configuração.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: o ranking é como a nota do jogo nos sites de crítica. Ajuda a escolher, mas você só sabe se gostou depois de jogar a fase que importa para você.

Para quem tem 65 anos: é como contratar um profissional. O diploma e as recomendações contam, mas a decisão vem depois de ver o trabalho feito em um caso real, com prazo, orçamento e clientes de verdade.

O QUE AINDA NÃO SABEMOS

Como medir de forma confiável agentes que trabalham por horas ou dias, em tarefas abertas. Se as provas privadas vão resistir à pressão de empresas que treinam para os rankings. Quanto da preferência humana é qualidade e quanto é estilo. E quem, no futuro, terá credibilidade para certificar que um modelo é seguro e adequado para áreas como saúde, finanças e educação.

O Teste Ana não é perfeito. É pequeno, depende do julgamento de quem aplica e não substitui avaliações técnicas sérias. Mas tem uma virtude que nenhum ranking global tem: ele mede o que importa para você. E, quando repetido de tempos em tempos, mostra se a ferramenta que você escolheu continua merecendo a sua confiança.

Isso nos leva ao último capítulo desta parte. Se os nomes mudam, os modelos mudam e até os termômetros mudam, como escrever, e ler, um livro sobre IA que não envelheça na semana seguinte? A resposta é separar o que dura do que muda: modelos mentais de um lado, fotografias datadas do outro. É o que chamo de Atlas Vivo das Novas Mentes.

O que mudou / O que significa / O que fazer agora

O que mudou: os rankings de IA se multiplicaram, mudam de versão e de vencedor com frequência, e já houve casos documentados de versões avaliadas diferentes das versões entregues ao público. Ao mesmo tempo, avaliações independentes passaram a usar provas privadas e tarefas mais próximas do trabalho real.

O que significa: um ranking mede um recorte. Alegação da empresa, avaliação independente e preferência do público são informações diferentes, e nenhuma delas substitui o teste no seu próprio trabalho, com seus dados, seu orçamento e suas exigências de confiança.

O que fazer agora: use os rankings para escolher os candidatos e o seu próprio teste para escolher o vencedor. Registre versão, data, preço e configuração de cada comparação. O melhor modelo do mundo para outra pessoa pode não ser o melhor para o seu trabalho.

AÇÃO EM DEZ MINUTOS

Escolha duas das oito tarefas do Teste Ana que mais se parecem com a sua semana. Entregue exatamente o mesmo pedido a duas IAs diferentes. Anote no seu Diário de Fronteira: data, ferramenta, modelo e versão, e uma nota de 1 a 5 em precisão, clareza e utilidade. O teste completo está no Apêndice B.

Notas

1. Zhang, Hugh et al. (Scale AI). "A Careful Examination of Large Language Model Performance on Grade School Arithmetic." NeurIPS 2024, Datasets and Benchmarks Track. https://proceedings.neurips.cc/paper_files/paper/2024/hash/53384f2090c6a5cac952c598fd67992f-Abstract-Datasets_and_Benchmarks_Track.html (uma versão anterior no arXiv relatava quedas de até 13%: https://arxiv.org/abs/2405.00332)

2. Stanford HAI. "Inside the AI Index: 12 Takeaways from the 2026 Report." 2026. https://hai.stanford.edu/news/inside-the-ai-index-12-takeaways-from-the-2026-report

3. Artificial Analysis. "Artificial Analysis Intelligence Index v4.1: a shift toward agentic workloads." 2026.

https://artificialanalysis.ai/articles/artificial-analysis-intelligence-index-v4-1; "Announcing Artificial Analysis Intelligence Index v4.2." 4 set. 2026. https://artificialanalysis.ai/articles/artificial-analysis-intelligence-index-v4-2

4. Chiang, Wei-Lin et al. "Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference." ICML 2024 (arXiv, 7 mar. 2024). https://arxiv.org/abs/2403.04132

5. LMArena (@lmarena_ai). Declaração sobre o Llama 4 no Arena, X, 8 abr. 2025. https://x.com/lmarena_ai/status/1909397817434816562

6. The Verge. "Meta gets caught gaming AI benchmarks with Llama 4." 7 abr. 2025 (atualizado). https://www.theverge.com/meta/645012/meta-llama-4-maverick-benchmarks-gaming; The Register, "Meta accused of Llama 4 bait-n-switch to juice LMArena rank." 8 abr. 2025. https://www.theregister.com/software/2025/04/08/meta-accused-of-llama-4-bait-n-switch-to-juice-lmarena-rank/1470166

7. Singh, Shivalika et al. "The Leaderboard Illusion." arXiv, abr. 2025; NeurIPS 2025, Datasets and Benchmarks Track. https://arxiv.org/abs/2504.20879

8. Artificial Analysis. "Artificial Analysis Intelligence Index" (v4.3.2) e "Announcing the Artificial Analysis Intelligence Index v4.3" (7 set. 2026) e "LLM Leaderboard". Consultados em 6 out. 2026. https://artificialanalysis.ai/evaluations/artificial-analysis-intelligence-index e https://artificialanalysis.ai/articles/artificial-analysis-intelligence-index-v4-3 e https://artificialanalysis.ai/leaderboards/models

9. Mitchell, Margaret et al. "Model Cards for Model Reporting." Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT*), 2019. https://arxiv.org/abs/1810.03993

10. OpenAI. "gpt-oss-120b & gpt-oss-20b Model Card." 5 ago. 2025. https://openai.com/index/gpt-oss-model-card/

11. Artificial Analysis. "Four frontier launches in eight days: six labs now field a model above 50 on the Artificial Analysis Intelligence Index." 17 jul. 2026. https://artificialanalysis.ai/articles/four-frontier-launches-in-eight-days-six-

labs-now-field-a-model-above-50-on-the-artificial-analysis-intelligence-index

CAPÍTULO 8

O Atlas Vivo das Novas Mentes

Tese do capítulo: o livro precisa ensinar modelos mentais duráveis e, ao mesmo tempo, oferecer uma fotografia atualizável do mercado.

Terminei o capítulo anterior com uma pergunta incômoda para quem escreve sobre IA: se os nomes mudam, os modelos mudam e até os termômetros mudam, como fazer um livro que não envelheça na semana seguinte?

Eu sinto essa pergunta na pele. Enquanto escrevia os capítulos anteriores, usei fotografias de julho de 2026. Quando cheguei a este, em outubro, quase todas já tinham sido substituídas. A OpenAI tinha lançado o GPT-6 Astra e, depois, o GPT-6.1 Sol. A Anthropic tinha lançado o Claude Opus 5.5. O Google tinha anunciado o Gemini 4 Argon.[1,2,3] E o Kimi K3, que em julho aparecia com 57 pontos no índice da Artificial Analysis, aparecia agora com 44, não porque tivesse piorado, mas porque o índice mudou de versão.[4,5]

Foi aí que entendi o que eu realmente queria oferecer. Não uma lista das minhas ferramentas favoritas. Preferência pessoal todo mundo tem, e a minha não vale mais do que a sua. O que eu queria era uma visão geral honesta: quem são os principais construtores, o que cada um tenta ser, o que mudou por último e como ler tudo isso sem se perder.

Chamo isso de Atlas Vivo das Novas Mentes. Um atlas, porque é um mapa. Vivo, porque precisa ser redesenhado o tempo todo. A parte impressa deste capítulo ensina a ler o mapa. A parte viva, que explico no fim, mantém o mapa atualizado.

Uma regra de leitura antes de começar: o que vem a seguir mistura **fatos datados**, ligados às notas, com a **minha leitura** de cada empresa, que é interpretação e deve ser tratada como tal.

1. Estados Unidos e Europa: OpenAI, Anthropic, Google, Microsoft, Meta, xAI, Mistral e Perplexity

OpenAI. A ambição, na minha leitura, é ser o assistente e a plataforma generalista: conversa, pesquisa, código, criação e agentes, com enorme distribuição. A fraqueza é o reverso disso: muitos modos, muitos produtos e uma pressão constante sobre custo, confiança e governança. Última mudança registrada: o GPT-6 Astra, em setembro de 2026, com preço de API de 10 dólares por milhão de tokens de entrada e 50 por milhão de saída, seguido no fim do mês pelo GPT-6.1 Sol.[1]

Anthropic. Aposta em trabalho intelectual profundo, código, documentos longos e agentes de longa duração, com forte ênfase pública em segurança. A fraqueza é uma distribuição menor ao consumidor e um comportamento mais cauteloso em alguns contextos. Última mudança: o Claude Opus 5.5, em 22 de setembro de 2026. Em 6 de outubro, ele liderava o índice de inteligência da Artificial Analysis.[2,5]

Google. É a única dessas empresas que já controla, ao mesmo tempo, Busca, Android, Gmail, YouTube e uma nuvem própria. A força é integrar inteligência a tudo isso, com multimodalidade,

ciência e escala. A fraqueza, na minha leitura, é uma experiência fragmentada e uma liderança desigual entre categorias. Última mudança: o Gemini 4 Argon, anunciado em 30 de setembro de 2026, inicialmente liberado apenas para um grupo de especialistas em defesa cibernética, antes de chegar a desenvolvedores, empresas e consumidores.[3]

Microsoft. Durante anos foi, sobretudo, a grande distribuidora de modelos de outros, com o Copilot movido por modelos da OpenAI e, mais tarde, também da Anthropic. Em junho de 2026, a Microsoft AI anunciou uma família de sete modelos próprios, a família MAI, com o MAI-Thinking-1 como modelo de raciocínio principal.[6] A pergunta a acompanhar: quanto do Copilot será movido por modelos da casa.

Meta. Popularizou os pesos abertos com a família Llama e hoje concentra seus modelos mais avançados na linha Muse, proprietária. A força é a distribuição nos aplicativos da própria empresa, como Facebook, Instagram e WhatsApp. Última mudança: o Muse Spark 1.3, em 2 de setembro de 2026, e um agente pessoal chamado Muse, apresentado no mesmo mês.[7]

xAI. Criadora do Grok e, desde fevereiro de 2026, parte da SpaceX, que hoje assina seus produtos como SpaceXAI.[8] Aposta em velocidade de lançamento e em preço competitivo. Última mudança: o Grok 4.7, em setembro de 2026, anunciado pela empresa como seu modelo mais capaz para programação e trabalho intelectual, a 2 dólares por milhão de tokens de entrada e 6 por milhão de saída.[9]

Mistral. A principal construtora europeia, com sede na França. Aposta em modelos eficientes, muitos com pesos abertos sob licença Apache 2.0, e em implantação dentro da infraestrutura do próprio cliente, um argumento forte para soberania e privacidade.

O Mistral Small 4, de março de 2026, é um exemplo.[10] A fraqueza: no índice geral de inteligência, seus modelos estão longe do topo.[5] Isso não os torna inúteis; os torna adequados a outras tarefas.

Perplexity. É, antes de tudo, uma camada de produto: um mecanismo de respostas com fontes e, desde março de 2026, o navegador Comet disponível em iOS, Android, Mac e Windows.[11] Usa modelos de outras empresas: em outubro de 2026, por exemplo, oferecia o GPT-6.1 Sol, da OpenAI, aos assinantes pagos.[11] A força é a experiência de pesquisa. A dependência de modelos de terceiros é, ao mesmo tempo, flexibilidade e vulnerabilidade.

2. China: DeepSeek, Qwen, Kimi, GLM, Seed, Hunyuan e ERNIE

O ecossistema chinês merece atenção por um motivo que vimos no Capítulo 6: boa parte da sua força chega ao mundo em forma de pesos abertos. E por outro motivo que todo leitor deveria considerar: privacidade e jurisdição.

DeepSeek. Ficou famosa pela eficiência e por publicar pesos sob licença MIT. Última mudança: o DeepSeek-V4.1-Flash, em 10 de setembro de 2026.[12] Atenção à privacidade: a política da empresa informa que dados pessoais podem ser armazenados em servidores na República Popular da China, e em janeiro de 2025 a autoridade italiana de proteção de dados bloqueou o tratamento de dados de usuários italianos pelo aplicativo.[13] Usar os pesos abertos nos seus próprios servidores é uma situação muito diferente de usar o aplicativo da empresa.

Qwen. A família de modelos da Alibaba, com muitas versões abertas e uma presença enorme entre desenvolvedores. Última

mudança: o Qwen3.8-Max, em agosto de 2026; a Alibaba já anunciou a família Qwen4, ainda sem data de lançamento.[14,15]

Kimi. Da Moonshot AI. Em julho de 2026, o Kimi K3 chegou ao terceiro lugar do índice da Artificial Analysis, e seus pesos foram publicados poucos dias depois, sob licença própria.[4,16] Em outubro, ainda estava entre os três modelos abertos mais bem colocados do índice.[5]

GLM. Da Z.ai. O GLM-5.3, de agosto de 2026, teve os pesos abertos. Em setembro, o centro de avaliação de IA do NIST, órgão do governo americano, concluiu que ele era o modelo open weight com maior capacidade cibernética já publicado, mas cerca de quatro meses atrás da fronteira americana nesse quesito.[17] É um bom exemplo de como capacidade aberta também levanta perguntas de segurança.

Seed. A equipe de modelos da ByteDance, dona do TikTok. A série Seed2.0, lançada em fevereiro de 2026, alimenta o aplicativo Doubao, que, segundo a empresa, atende centenas de milhões de usuários. O acesso para empresas é via API, na plataforma Volcano Engine.[18]

Hunyuan. Da Tencent. O Hy4 preview, de 28 de agosto de 2026, foi publicado com pesos abertos sob licença Apache 2.0 e integrado a produtos da empresa.[19]

ERNIE. Do Baidu, uma das pioneiras chinesas em grandes modelos. Última mudança: o ERNIE 5.1, em maio de 2026, com foco declarado em busca, raciocínio e escrita.[20]

E o mapa nunca está completo. Em 6 de outubro de 2026, o modelo aberto mais bem colocado no índice da Artificial Analysis não era de nenhuma dessas sete famílias: era o MiMo-V2.6-Pro,

da Xiaomi.[5] Um atlas honesto precisa ter espaço para quem ainda não está nele.

FOTOGRAFIA DA FRONTEIRA — 6 DE OUTUBRO DE 2026

Artificial Analysis Intelligence Index v4.3.2: 1.º Claude Opus 5.5, 58 pontos. GPT-6 Astra e Gemini 4 Argon, 53. Muse Spark 1.3, 48. Grok 4.7, 46. Melhores modelos abertos: MiMo-V2.6-Pro, 46; GLM-5.3, 45; Kimi K3, 44. Em julho, com outra versão do índice, os números eram outros; não compare notas de versões diferentes.[5]

Esta fotografia já pode estar velha quando você a ler: confira a versão atualizada no Atlas Vivo da IA.

NÃO CONFUNDA

Ficar em primeiro lugar em um índice não significa ser a melhor empresa, o melhor produto ou a melhor escolha para você. E nacionalidade não é, sozinha, um critério de qualidade: um modelo chinês rodando nos seus servidores e um aplicativo chinês recebendo seus dados são situações diferentes, assim como um modelo americano pode ter regras de dados inadequadas para o seu caso.

3. QR code e página viva: missão, força, fraqueza, privacidade, preço e última mudança

Se a fotografia envelhece, o que dura? Os critérios. Por isso, cada ficha do Atlas Vivo responde sempre às mesmas seis perguntas:

Missão. O que essa empresa está tentando ser? Assistente universal, ferramenta de trabalho profundo, infraestrutura, plataforma aberta?

Força. Em que ela é comprovadamente boa, segundo fontes primárias e avaliações independentes, e não segundo a propaganda?

Fraqueza. O que ela ainda faz mal, ou o que você abre mão ao escolhê-la?

Privacidade. Onde ficam os dados, se são usados para treinamento, se existe opção de não retenção, qual lei se aplica. A OpenAI e a Anthropic, por exemplo, informam oferecer retenção zero de dados para clientes elegíveis da API em seus modelos mais recentes.[1,2] É o tipo de detalhe que muda a decisão de uma empresa.

Preço. Quanto custa para uma pessoa e quanto custa por milhão de tokens na API, lembrando que preço por token e custo por tarefa são coisas diferentes.

Última mudança. O que aconteceu de mais recente, com data e fonte.

Na edição impressa, um QR code levará à página viva do Atlas. Ali, as fichas serão revisadas seguindo o mesmo calendário editorial que adotei para todo o livro: fatos vivos são conferidos 30 dias antes do fechamento do manuscrito, de novo sete dias antes da impressão e, depois, continuamente na versão digital. Entre uma revisão e outra, a página registra o que mudou, para que você possa acompanhar a história, e não só o placar.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: é como a página de um jogador em um app de futebol. A posição e o estilo de jogo mudam pouco; as estatísticas mudam a cada rodada. Você precisa das duas coisas para escalar o time.

Para quem tem 65 anos: é como um guia de viagem com um encarte de atualizações. O guia ensina a conhecer a cidade; o encarte avisa que aquele restaurante fechou e que a ponte está em obras.

O QUE AINDA NÃO SABEMOS

Quantos desses nomes ainda estarão na fronteira daqui a dois anos (horizonte próximo, confiança baixa). Se os modelos abertos voltarão a encostar nos fechados. Se novos países e novas empresas entrarão no mapa. E se os próprios índices conseguirão acompanhar a velocidade dos lançamentos sem mudar de régua a cada trimestre.

Volto, então, à pergunta do começo. Este livro não vai envelhecer na semana seguinte se fizer duas coisas ao mesmo tempo: ensinar modelos mentais que duram, como as camadas do Capítulo 5, os dois eixos do Capítulo 6 e o Teste Ana do Capítulo 7, e manter uma fotografia datada, que qualquer um pode conferir e atualizar.

O que mudou / O que significa / O que fazer agora

O que mudou: em poucas semanas, entre setembro e outubro de 2026, os principais laboratórios lançaram novas gerações de modelos, e o principal índice independente mudou de líder e de versão. Ao mesmo tempo, o mapa ficou mais diverso: Estados

Unidos, Europa e China, laboratórios fechados e abertos, empresas de modelo e empresas de produto.

O que significa: qualquer lista de "melhores IAs" é uma fotografia com prazo de validade. O que permanece são as perguntas: missão, força, fraqueza, privacidade, preço e última mudança.

O que fazer agora: antes de adotar uma ferramenta, monte a ficha dela com essas seis perguntas e anote a data. Revise a ficha uma vez por mês. O atlas muda; os critérios de leitura permanecem.

AÇÃO EM DEZ MINUTOS

Escolha a IA que você mais usa e preencha, no seu Diário de Fronteira, a ficha de seis linhas: missão, força, fraqueza, privacidade, preço e última mudança, com a data de hoje. Na linha de privacidade, procure na política da empresa se as suas conversas podem ser usadas para treinamento e como desativar isso.

Fim da Parte II

Chegamos ao fim da segunda parte. Aprendemos a separar empresa, laboratório, modelo, produto e agente. Vimos que aberto e frontier são eixos diferentes. Descobrimos que todo ranking mede um recorte e que o melhor teste é o seu. E, agora, temos um mapa e um método para mantê-lo atualizado.

Saber o que é o quê era o segundo passo. O terceiro é entender a disputa.

Porque esse mapa não é uma vitrine tranquila. É uma corrida. Empresas com histórias, missões e formas de distribuição muito diferentes estão disputando a mesma fronteira, com bilhões de

dólares, os melhores talentos do mundo e, cada vez mais, o interesse de governos inteiros. A OpenAI quer ser o assistente de todos. A Anthropic aposta no trabalho profundo e confiável. O Google tenta espalhar inteligência por tudo o que já tem. E, do outro lado do mundo, a China corre com outra lógica.

Na Parte III, vamos entrar nessa corrida. Não para escolher um campeão, mas para entender o que cada competidor está realmente tentando construir, e o que isso significa para você.

Notas

1. OpenAI. "GPT-6 Astra: A new generation of intelligence." Set. 2026 (atualizado em 22 e 29 set. 2026). https://openai.com/index/gpt-6-astra/; OpenAI, "Introducing GPT-6.1 Sol." 29 set. 2026. https://openai.com/index/introducing-gpt-6-1-sol; France 24, "OpenAI begins rollout of GPT-6 amid growing scrutiny over safety." 4 set. 2026. https://www.france24.com/en/technology/20260904-openai-begins-rollout-of-gpt-6-with-focus-on-cyber-security-safeguards

2. Anthropic. "Introducing Claude Opus 5.5." 22 set. 2026. https://www.anthropic.com/claude-opus-5-5

3. Kavukcuoglu, Koray. "Gemini 4 Argon: our next era of frontier intelligence." Google, 30 set. 2026. https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-4-argon/

4. Artificial Analysis. "Four frontier launches in eight days: six labs now field a model above 50 on the Artificial Analysis Intelligence Index." 17 jul. 2026. https://artificialanalysis.ai/articles/four-frontier-launches-in-eight-days-six-labs-now-field-a-model-above-50-on-the-artificial-analysis-intelligence-index

5. Artificial Analysis. "LLM Leaderboard" (Intelligence Index v4.3.2). Consultado em 6 out. 2026. https://artificialanalysis.ai/leaderboards/models

6. Suleyman, Mustafa. "Building a hill-climbing machine: Launching seven new MAI models." Microsoft AI, 2 jun. 2026. https://microsoft.ai/news/building-a-hillclimbing-machine-launching-seven-new-mai-models/

7. Meta. "Introducing Muse Spark 1.3." Meta AI Research, 2 set. 2026. https://research.meta.ai/blog/introducing-muse-spark-1-3; Meta, "Introducing Muse: The World's First Personal AI Agent Built for Everyone." Set. 2026. https://about.fb.com/news/2026/09/introducing-muse-personal-ai-agent/

8. xAI. "xAI joins SpaceX." 2 fev. 2026. https://x.ai/news/xai-joins-spacex

9. SpaceXAI. "Introducing Grok 4.7." Set. 2026. https://x.ai/news/grok-4-7

10. Mistral AI. "Introducing Mistral Small 4." 16 mar. 2026. https://mistral.ai/news/mistral-small-4/

11. Perplexity. "What we shipped — March 27, 2026" (Comet iOS). https://www.perplexity.ai/changelog/what-we-shipped--march-27-2026; Perplexity, "Comet." https://www.perplexity.ai/comet; Perplexity, changelog "Automations, inline visualizations and GPT-6.1 Sol." Out. 2026. https://www.perplexity.ai/changelog/automations-inline-visualizations-and-gpt-6-1-sol

12. DeepSeek. "Introducing DeepSeek-V4.1-Flash: smarter, faster, more efficient." 10 set. 2026. https://www.deepseek.com/en/news/deepseek-v4-1-flash/

13. Garante per la protezione dei dati personali. Provvedimento del 30 gennaio 2025 [10098477]. https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/10098477; Reuters, "Italy's regulator blocks Chinese AI app DeepSeek on data protection." 30 jan. 2025. https://www.reuters.com/technology/artificial-intelligence/italys-privacy-watchdog-blocks-chinese-ai-app-deepseek-2025-01-30/

14. Qwen. "Qwen3.8-Max: A New Bar for Coding and Cowork." 2 ago. 2026. https://qwen.ai/blog?id=qwen3.8

15. Alibaba Cloud. "Alibaba Unveils Roadmap on Full-Stack AI Strategy from Chips, Cloud Infrastructure, Models to Agents." 2026.

https://www.alibabacloud.com/en/press-room/alibaba-unveils-roadmap-on-full-stack-ai-strategy

16. Moonshot AI. "Kimi-K3" (README e licença). Hugging Face, 27 jul. 2026. https://huggingface.co/moonshotai/Kimi-K3

17. NIST, Center for AI Standards and Innovation (CAISI). "CAISI's Assessment of Z.ai's GLM-5.3 Cyber Capabilities." 17 set. 2026. https://www.nist.gov/news-events/news/2026/09/caisis-assessment-zais-glm-53-cyber-capabilities

18. ByteDance Seed. "Seed 2.0 Official Launch." 14 fev. 2026. https://seed.bytedance.com/en/blog/seed-2-0-official-launch

19. Tencent. "Tencent Releases and Open-Sources Tencent Hy4 preview." 28 ago. 2026. https://www.tencent.com/tencent-releases-and-open-sources-tencent-hy4-preview/

20. Baidu. "ERNIE 5.1 Officially Released!" ERNIE Blog, maio 2026. https://ernie.baidu.com/blog/posts/ernie-5.1-0508-release/

PARTE III

A corrida pela fronteira

CAPÍTULO 9

OpenAI, Anthropic e Google: três ideias de futuro

Tese do capítulo: as três empresas disputam a fronteira, mas sua história, missão, distribuição e produto levam a estratégias diferentes.

Terminei a segunda parte com uma frase curta: a OpenAI quer ser o assistente de todos, a Anthropic aposta no trabalho profundo e confiável, e o Google tenta espalhar inteligência por tudo o que já tem. São simplificações. Mas foi assim que comecei a entender a corrida: não pela pergunta "quem tem o melhor modelo?", e sim pela pergunta "o que cada empresa está tentando construir?".

Percebi isso no meu próprio trabalho. Não uso uma única IA. Uso uma para conversar e organizar ideias, outra para ler documentos longos e revisar textos com cuidado, outra dentro das ferramentas que já fazem parte do meu dia. Uso IAs americanas e muitas chinesas, e tenho agentes separados para escrever livros, construir sites, cuidar dos e-mails, organizar o imobiliário, pesquisar ações e estudar humanoides.

No varejo, eu aprendi que duas empresas podem vender o mesmo tipo de produto e, ainda assim, apostar em futuros completamente diferentes. Uma quer estar em todas as prateleiras. Outra quer ser a escolha do cliente exigente. Outra já é dona do shopping inteiro. No Grupo Pão de Açúcar, eu me sentava à mesa com os grandes fabricantes da época, Compaq, HP e IBM,

além de Microsoft, Intel e AMD, para negociar contratos de fornecimento, justamente quando abríamos o supermercado como canal de venda de computadores. Depois, mudei de lado do balcão: na IBM Brasil, como Diretora Nacional de Vendas da linha PSG, de computadores pessoais, para o varejo, liderei uma divisão de 280 milhões de dólares. Vi a mesma disputa pelos dois lados da mesa.

Com a inteligência artificial, acontece algo parecido. Neste capítulo, olho para as três empresas americanas mais observadas da fronteira com a mesma régua: de onde vêm, o que dizem querer e como chegam até você. O que segue mistura **fatos datados**, ligados às notas, com a **minha leitura**, que é interpretação.

1. ChatGPT: o assistente que quer ser plataforma

A OpenAI nasceu com uma missão ampla. Sua carta de princípios, publicada em 2018, diz que a missão é garantir que a inteligência artificial geral, a AGI, beneficie toda a humanidade.[1] Uma missão para "toda a humanidade" combina com um produto para todo mundo. E o ChatGPT virou exatamente isso.

Em 29 de setembro de 2026, na sua conferência anual para desenvolvedores, a OpenAI informou que o ChatGPT tinha mais de 1,2 bilhão de usuários semanais, contra 1 bilhão em julho.[2] É um número da própria empresa, mas dá a dimensão da aposta: o ChatGPT quer ser a primeira porta que alguém abre diante de uma pergunta, uma tarefa ou uma ideia.

A ambição, porém, não para na conversa. Na mesma conferência, segundo a cobertura do evento, a OpenAI apresentou os Dots, agentes que ficam sempre ligados, têm o próprio computador na nuvem, se conectam a mais de 4 mil aplicativos e

podem ser acionados pelo ChatGPT, pelo Slack ou pelo Microsoft Teams. Anunciou também um login com a conta do ChatGPT para outros aplicativos e um marketplace para desenvolvedores venderem produtos a empresas.[3] Quando outros negócios constroem em cima de você, você deixa de ser produto e vira plataforma.

Os modelos acompanham essa lógica de escada. No topo, o GPT-6 Astra, lançado no início de setembro de 2026, custa 10 dólares por milhão de tokens de entrada e 50 por milhão de saída. Logo abaixo, o GPT-6.1 Sol, de 29 de setembro, custa 2 e 10 dólares, e a OpenAI o apresenta como uma inteligência próxima à do Astra por um quinto do preço.[4] Um modelo de luxo e um de volume: a estratégia de quem quer atender do cliente mais exigente ao mais sensível a preço.

E há um terceiro sinal, talvez o mais revelador. Em 5 de outubro de 2026, a OpenAI anunciou um novo formato de anúncio visual no ChatGPT, a ser testado nos Estados Unidos durante a geração de imagens. Os anúncios já existiam no produto desde o início do ano, nos planos gratuitos e mais baratos; segundo a empresa, eles não influenciam as respostas, e assinantes Plus e Pro e clientes empresariais não os veem.[5] Quem vende assinaturas, API e publicidade está construindo um negócio de massa, como os grandes negócios de internet que já conhecemos.

Na minha leitura, a força da OpenAI é a amplitude: pesquisa, código, criação, agentes e uma distribuição enorme. A fraqueza é o reverso da mesma moeda. São muitos modos, muitos planos e muitos produtos, e cada nova frente aumenta a pressão sobre custo, confiança e governança. Quando o mesmo assistente conversa com você, trabalha nas suas ferramentas e mostra anúncios, a pergunta "a quem ele serve?" passa a merecer atenção.

NÃO CONFUNDA

Usuários semanais não são usuários mensais. Em 2026, a OpenAI divulgou usuários semanais do ChatGPT, e o Google, usuários mensais do aplicativo Gemini. Os dois números não se comparam diretamente. E nenhum deles mede receita, lucro ou qualidade: mede alcance, segundo a própria empresa.

2. Claude: a aposta no trabalho profundo

A Anthropic conta outra história. Foi fundada em 2021 por ex-integrantes da OpenAI e se organiza como empresa de benefício público, com o propósito declarado de desenvolver e manter IA avançada de forma responsável, para o benefício de longo prazo da humanidade.[6] Na prática, ela disputa a fronteira por um caminho mais estreito e mais fundo: o trabalho profissional.

O caminho começou pelo código. O Claude Code levou o modelo para dentro do trabalho dos programadores, não como um conselheiro, mas como alguém que recebe a tarefa inteira. Em janeiro de 2026, a empresa lançou o Claude Cowork, descrito por ela como "o Claude Code para o resto do seu trabalho": um agente que lê, organiza e cria arquivos no computador de quem não é programador. Começou como prévia de pesquisa para assinantes do plano Max e, ao longo do ano, chegou aos planos pagos com controles para empresas.[7]

O Claude Opus 5.5, lançado em 22 de setembro de 2026, mostra bem a aposta. A Anthropic o apresentou com exemplos de trabalho longo: segundo a empresa, um cliente de testes concluiu em menos de um dia uma migração de 680 mil linhas de código, e outro deixou o modelo trabalhando sozinho durante a noite, por mais de 18 horas, em seis repositórios. O preço caiu para 4 dólares

por milhão de tokens de entrada e 20 por milhão de saída, 20% menos que o do Opus 5, e o modelo está disponível nas nuvens da Amazon, do Google e da Microsoft, com opção de retenção zero de dados.[8] São alegações da empresa. Mas, em 6 de outubro de 2026, o Opus 5.5 também liderava o índice de inteligência da Artificial Analysis, uma avaliação independente, com 58 pontos.[9]

A outra metade da identidade da Anthropic é a segurança, e ela tem consequências comerciais visíveis. Em setembro de 2026, o diretor-executivo da empresa, Dario Amodei, publicou um ensaio defendendo que a indústria desacelere de forma deliberada o ritmo de avanço das capacidades, para que as práticas de segurança acompanhem, e comprometeu a Anthropic a receber avaliadores externos com acesso semelhante ao de funcionários.[10] A empresa também mantém o Claude Mythos 5.1 restrito a programas de acesso confiável, e o próprio Opus 5.5 chegou com salvaguardas que, segundo a Anthropic, redirecionam a maior parte das tarefas de cibersegurança para um modelo anterior.[8,11]

Na minha leitura, a força da Anthropic é a profundidade: código, escrita, documentos longos e tarefas que duram horas. A fraqueza é uma distribuição menor ao consumidor e um comportamento mais cauteloso em alguns contextos. Para quem põe um agente para trabalhar sozinho dentro da empresa, essa cautela pode ser o argumento de venda; para quem quer liberdade máxima, um obstáculo.

Não existe um campeão universal. Existe uma missão, e uma ferramenta mais adequada a ela.

3. Gemini: a inteligência distribuída por tudo

O Google entra na corrida de um lugar que nenhuma das outras duas empresas ocupa. A missão do Google DeepMind, o laboratório da empresa, é construir IA de forma responsável para beneficiar a humanidade; a missão histórica do Google, lembrada pelo próprio laboratório, é organizar a informação do mundo e torná-la universalmente acessível e útil.[12] A diferença é que o Google já tem onde colocar essa inteligência.

Em 11 de agosto de 2026, o Google informou que o aplicativo Gemini tinha passado de 1 bilhão de usuários mensais, o produto de crescimento mais rápido da história da empresa. No Android, segundo o Google, o Gemini já automatizava ações em mais de 40 aplicativos populares; no iPhone, tinha mais de 100 milhões de usuários ativos.[13] No resultado do segundo trimestre de 2026, a Alphabet relatou que quase 90% das cem maiores empresas da lista Fortune usavam o Gemini Enterprise, que seus modelos processavam 22 bilhões de tokens por minuto via API e que os recursos de IA estavam impulsionando o crescimento das buscas.[14] E o Gemini está dentro do Gmail, do Docs, do Sheets e do Meet, para quem usa o Google Workspace.[15]

A estratégia vai além do escritório. Na ciência, o AlphaFold, criado pelo DeepMind para prever a estrutura de proteínas, rendeu a Demis Hassabis e John Jumper metade do Prêmio Nobel de Química de 2024.[16] Na robótica, o Google DeepMind apresentou em 30 de julho de 2026 o Gemini Robotics 2, um modelo que, segundo a empresa, controla robôs humanoides inteiros, dos pés às pontas dos dedos; por enquanto, a versão que controla os movimentos está disponível apenas para parceiros de acesso antecipado.[17]

E a fronteira? Em 30 de setembro de 2026, o Google anunciou o Gemini 4 Argon, liberado inicialmente para um grupo de especialistas em defesa cibernética; segundo a empresa, ele chegará depois a desenvolvedores, empresas e consumidores, começando por clientes pagos da API e assinantes do Google AI Ultra. O preço introdutório é de 2 dólares por milhão de tokens de entrada e 10 por milhão de saída, subindo para 4 e 20 depois desse período; o limite de saída foi ampliado para 1 milhão de tokens.[18] No índice da Artificial Analysis de 6 de outubro, o Argon empatava com o GPT-6 Astra, com 53 pontos.[9]

Na minha leitura, a força do Google é a escala: multimodalidade, custo, ciência, produtos e distribuição. A fraqueza é uma experiência fragmentada, com o Gemini aparecendo em lugares e versões diferentes, e uma liderança desigual entre categorias. Ser dono do shopping é uma vantagem enorme, mas não garante a preferência do cliente.

FOTOGRAFIA DA FRONTEIRA — TRÊS APOSTAS (6 DE OUTUBRO DE 2026)

OpenAI: assistente e plataforma generalista. ChatGPT com mais de 1,2 bilhão de usuários semanais (29 set.); GPT-6 Astra (US$ 10/50) e GPT-6.1 Sol (US$ 2/10); agentes Dots; anúncios nos planos gratuitos e mais baratos.

Anthropic: trabalho profundo e agentes de longa duração. Claude Opus 5.5 (22 set., US$ 4/20), 1.º no índice da Artificial Analysis, com 58 pontos; Claude Code e Cowork; Mythos 5.1 restrito.

Google: inteligência distribuída. Aplicativo Gemini com mais de 1 bilhão de usuários mensais (ago.); Gemini 4 Argon (30 set., acesso inicial restrito, US$ 2/10 introdutório), com 53 pontos; Gemini Robotics 2 (30 jul.).

Preços por milhão de tokens de entrada/saída. Fatos vivos: confira o Atlas Vivo da IA.

O que isso significa para você

Se cada empresa aposta em um futuro diferente, a pergunta útil deixa de ser "qual é a melhor IA?" e passa a ser "qual dessas apostas combina com o meu trabalho?".

Se você precisa de um assistente versátil, que converse, pesquise, crie imagens e se conecte a muitos aplicativos, a aposta da OpenAI foi desenhada para você. Mas vale ler as regras do plano que você usa, inclusive sobre anúncios e dados. Se o seu trabalho depende de documentos longos, código, análise cuidadosa ou agentes que trabalham por horas, a aposta da Anthropic foi desenhada para isso, e você deve esperar mais cautela em alguns temas. Se a sua vida e a sua empresa já rodam no Gmail, no Docs, no Android e na nuvem do Google, a

inteligência vai chegar até você por ali, quer você escolha ativamente ou não.

Há também uma questão de poder. Cada arquitetura cria uma dependência diferente. Com a OpenAI, você depende de um assistente que quer estar no centro de tudo. Com a Anthropic, de um parceiro de trabalho especializado, que muitas vezes chega a você por meio da nuvem ou do produto de outras empresas. Com o Google, de um ecossistema inteiro. Nenhuma dependência é boa ou ruim por si só. Mas um executivo que não sabe de quem depende não está decidindo; está sendo decidido.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: pense em três jogos de sucesso. Um quer ser o jogo que todo mundo joga. Outro é o preferido de quem leva o jogo a sério. O terceiro já vem instalado no seu celular. Os três podem ser ótimos; a escolha depende de como você joga.

Para quem tem 65 anos: é como escolher entre a grande loja de departamentos, o especialista de confiança e o banco onde você já tem conta, cartão e seguro. Cada um ganha a sua preferência por um motivo diferente.

O QUE AINDA NÃO SABEMOS

Se a publicidade vai mudar a relação de confiança entre as pessoas e o ChatGPT. Se a defesa da Anthropic por um ritmo mais lento vai ser seguida pelos concorrentes ou vai custar terreno na corrida. Se o Google vai conseguir transformar distribuição em preferência. E se, daqui a dois anos, essas três estratégias ainda serão tão diferentes ou vão convergir para o mesmo produto.

O que mudou / O que significa / O que fazer agora

O que mudou: a disputa pela fronteira deixou de ser apenas uma corrida por modelos melhores. Em setembro e outubro de 2026, as três empresas lançaram novas gerações de modelos, mas também mostraram estratégias distintas: plataforma e publicidade na OpenAI, agentes de trabalho profundo e ritmo deliberado na Anthropic, distribuição por todo o ecossistema no Google.

O que significa: comparar ChatGPT, Claude e Gemini apenas por pontos em um ranking é olhar para a vitrine e ignorar a loja. Missão, distribuição, modelo de negócio e regras de segurança dizem tanto quanto o desempenho.

O que fazer agora: antes de escolher uma ferramenta, pergunte qual é a aposta da empresa que a constrói e se ela combina com o seu trabalho. Não existe um campeão universal; existe uma arquitetura de poder diferente em cada empresa.

AÇÃO EM DEZ MINUTOS

Escolha uma tarefa real desta semana e entregue o mesmo pedido ao ChatGPT, ao Claude e ao Gemini. Anote no seu Diário de Fronteira, com a data: qual resposta foi mais útil, qual exigiu menos correção e qual se encaixou melhor nas ferramentas que você já usa. Depois, escreva em uma linha qual das três apostas combina com a sua missão.

E aqui surge uma pergunta que muita gente me faz em voz baixa. Se o Google tem tudo isso, a Busca, o Android, a nuvem, o Nobel, os robôs, por que tanta gente fala como se ele estivesse atrás? É o assunto do próximo capítulo.

Notas

1. OpenAI. "OpenAI Charter." 9 abr. 2018. https://openai.com/charter/

2. Peters, Jay. "ChatGPT now has 1.2 billion weekly users, OpenAI says." The Verge, 29 set. 2026. https://www.theverge.com/ai-artificial-intelligence/1002265/chatgpt-now-has-1-2-billion-weekly-users-openai-says; The Decoder, "ChatGPT now reaches 1.2 billion people every week, OpenAI says." 29 set. 2026. https://the-decoder.com/chatgpt-now-reaches-1-2-billion-people-every-week-openai-says/

3. SN, Vikas. "OpenAI DevDay 2026: Dots always-on AI agents; GPT-6.1 Sol; ChatGPT subscriptions in 3rd party tools." Moneycontrol, 30 set. 2026. https://www.moneycontrol.com/artificial-intelligence/openai-devday-2026-dots-always-on-ai-agents-gpt-6-1-sol-chatgpt-subscriptions-in-3rd-party-tools-article-14041450.html

4. OpenAI. "GPT-6 Astra: A new generation of intelligence." Set. 2026 (atualizado em 22 e 29 set. 2026). https://openai.com/index/gpt-6-astra/; OpenAI API, "GPT-6.1 Sol Model." https://developers.openai.com/api/docs/models/gpt-6.1-sol; The Next Web, "OpenAI releases GPT-6.1 Sol at a fifth of GPT-6 Astra's token prices." 29 set. 2026. https://thenextweb.com/news/openai-gpt-6-1-sol-price-astra-devday

5. OpenAI. "Building advertising for the way people use AI." 5 out. 2026. https://openai.com/index/new-chatgpt-ads-format-and-measurement/; Bonifacic, Igor. "OpenAI is testing more ads in ChatGPT." Engadget, 5 out. 2026. https://www.engadget.com/2277271/openai-is-testing-more-ads-in-chatgpt/; Scanlon, Krystal. "OpenAI moves to make ChatGPT Ads more measurable — and more visual." Digiday, 5 out. 2026. https://digiday.com/marketing/openai-moves-to-make-chatgpt-ads-more-measurable-and-more-visual/

6. Anthropic. "Company" e "The Long-Term Benefit Trust." Consultados em 6 out. 2026. https://www.anthropic.com/company e https://www.anthropic.com/news/the-long-term-benefit-trust

7. Fortune. "Anthropic launches Claude Cowork, a file-managing AI agent that could threaten dozens of startups." 13 jan. 2026.

https://fortune.com/2026/01/13/anthropic-claude-cowork-ai-agent-file-managing-threaten-startups/; Anthropic, "Making Claude Cowork ready for enterprise." 2026. https://claude.com/blog/cowork-for-enterprise

8. Anthropic. "Introducing Claude Opus 5.5." 22 set. 2026. https://www.anthropic.com/claude-opus-5-5

9. Artificial Analysis. "LLM Leaderboard" (Intelligence Index v4.3.2). Consultado em 6 out. 2026. https://artificialanalysis.ai/leaderboards/models

10. Amodei, Dario. "We Must Pace the Frontier." Set. 2026. https://darioamodei.com/post/we-must-pace-the-frontier

11. Anthropic. "Introducing Claude Fable 5.1 and Claude Mythos 5.1." Set. 2026. https://www.anthropic.com/claude-fable-and-mythos-5-1

12. Google DeepMind. "About Google DeepMind." Consultado em 6 out. 2026. https://deepmind.google/about/

13. Google. "More than 1 billion people are using the Gemini app every month." 11 ago. 2026. https://blog.google/innovation-and-ai/products/gemini-app/one-billion-monthly-users/

14. Alphabet. "Alphabet Announces Second Quarter 2026 Results." 22 jul. 2026 (Exhibit 99.1, SEC). https://www.sec.gov/Archives/edgar/data/1652044/000165204426000066/googexhibit991q22026.htm

15. Google Workspace. "AI tools for business." Consultado em 6 out. 2026. https://workspace.google.com/solutions/ai/

16. The Royal Swedish Academy of Sciences. "The Nobel Prize in Chemistry 2024" (comunicado de imprensa). 9 out. 2024. https://www.nobelprize.org/prizes/chemistry/2024/press-release/

17. Parada, Carolina. "Gemini Robotics 2 brings whole body intelligence to robots." Google DeepMind, 30 jul. 2026. https://deepmind.google/blog/gemini-robotics-2-brings-whole-body-intelligence-to-robots/

18. Kavukcuoglu, Koray. "Gemini 4 Argon: our next era of frontier intelligence." Google, 30 set. 2026. https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-4-argon/

CAPÍTULO 10

O Google está ficando para trás?

Tese do capítulo: o Google é um competidor frontier, mas ser frontier não significa liderar toda categoria nem transformar capacidade técnica em preferência de usuário.

Terminei o capítulo anterior com a pergunta que muita gente me faz em voz baixa. Se o Google tem a Busca, o Android, a nuvem, o Nobel e os robôs, por que tanta gente fala como se ele estivesse atrás?

Para a minha geração, essa pergunta soa quase absurda. O Google foi, durante anos, a própria definição de futuro. Era a empresa que todo mundo queria imitar, o lugar onde os melhores engenheiros sonhavam trabalhar, a palavra que virou verbo.

Hoje, quando converso com jovens talentosos, ouço outros nomes. OpenAI, Anthropic, um laboratório novo, uma startup própria. O Google continua admirado, mas perdeu o monopólio simbólico de ser o lugar onde o futuro acontece. É uma mudança sutil, e mudanças sutis são justamente as que os líderes mais demoram a perceber.

Mas atenção: sensação não é fato. Este capítulo separa as duas coisas.

A desconfiança tem história. Em fevereiro de 2023, poucos meses depois do lançamento do ChatGPT, um anúncio do Google mostrou o seu novo chatbot, o Bard, dando uma resposta errada

sobre o telescópio James Webb. No mesmo dia, a Alphabet perdeu cerca de 100 bilhões de dólares em valor de mercado.[1] Ali nasceu a narrativa de que o Google tinha sido pego de surpresa. Três anos e meio depois, ela voltou.

1. Onde o Gemini está hoje

Comecemos pelo calendário, porque ele explica boa parte do ruído. Em 19 de maio de 2026, na sua conferência anual, o Google lançou o Gemini 3.5 Flash e disse que o 3.5 Pro, seu modelo mais avançado, já era usado internamente e chegaria "no próximo mês".[2] Não chegou. Em agosto, a imprensa especializada registrava que, meses depois, só tinham saído novos modelos Flash, mais leves.[3] Quando o Google finalmente anunciou o Gemini 4 Argon, em 30 de setembro, a Reuters informou que o 3.5 Pro não seria mais lançado e que o Argon, por enquanto, estava restrito a parceiros de cibersegurança, sem data para o público.[4]

Enquanto isso, OpenAI e Anthropic lançaram novas gerações em setembro. No índice da Artificial Analysis de 6 de outubro de 2026, o Claude Opus 5.5 liderava com 58 pontos; o Argon empatava com o GPT-6 Astra, com 53.[5] Ou seja: o Google está na fronteira, mas não no topo desse índice, e com o seu melhor modelo ainda fora do alcance da maioria das pessoas.

Agora olhe para as outras três dimensões, e o retrato muda.

Velocidade. O Gemini 3.6 Flash, lançado em julho de 2026, gerava cerca de 197 tokens por segundo na API do Google, segundo a Artificial Analysis, o que o colocava entre os 16 mais rápidos de 224 modelos comparáveis. A mesma avaliação, porém, registrava um tempo alto até a primeira palavra da resposta, de cerca de 15 segundos.[6] Rápido para produzir, lento para começar: a velocidade também tem mais de uma medida.

Multimodalidade. O mesmo modelo aceita texto, imagem, fala e vídeo e trabalha com uma janela de contexto de 1 milhão de tokens, o equivalente, segundo a Artificial Analysis, a cerca de 1.500 páginas.[6] Ler um vídeo, uma planilha e um contrato na mesma conversa é uma tradição da casa.

Custo. O 3.6 Flash custava 0,75 dólar por milhão de tokens de entrada e 3,75 por milhão de saída, contra medianas de 2 e 10 dólares entre modelos da sua faixa.[6] Não por acaso, segundo a Reuters, o Google passou nos últimos meses a destacar menos a liderança em capacidade e mais a vantagem de custo.[4]

E há ainda a preferência dos usuários, que é outra régua. No ranking do LMArena, baseado em votos de pessoas que comparam respostas às cegas, o Gemini 4 Argon aparecia em primeiro lugar em 2 de outubro de 2026, ainda com resultado preliminar e menos de 5 mil votos; o GPT-6 Astra estava em 29.º.[7] Dois rankings respeitados, duas fotografias diferentes do mesmo mercado. Como vimos no Capítulo 7, cada um mede um recorte.

FOTOGRAFIA DA FRONTEIRA — ONDE ESTÁ O GOOGLE (6 DE OUTUBRO DE 2026)

Capacidade: Gemini 4 Argon, 53 pontos na Artificial Analysis (Opus 5.5 lidera, com 58); acesso restrito, sem data pública. Gemini 3.5 Pro, prometido para junho, não lançado.

Preferência: Argon em 1.º no LMArena (preliminar, 2 out.).

Velocidade e custo: Gemini 3.6 Flash, cerca de 197 tokens/s; US$ 0,75/3,75 por milhão de tokens.

Uso na web: Gemini com 25,6% das visitas a sites de IA generativa em agosto de 2026, contra 55,5% do ChatGPT (Similarweb, só web).

Fatos vivos: confira o Atlas Vivo da IA.

2. As vantagens que não aparecem no ranking

Se o capítulo terminasse aqui, a resposta seria "um pouco atrás, mas perto". Só que a corrida não se decide apenas no modelo. Ela se decide também no que está em volta dele, e é aí que o Google tem vantagens que nenhum ranking mede.

Distribuição. No Capítulo 9, vimos o tamanho do aplicativo Gemini. Mas a vitrine mais poderosa do Google é a Busca. Em maio de 2026, o Google informou que o AI Mode, a busca em formato de conversa, tinha passado de 1 bilhão de usuários mensais e que os recursos de IA eram o principal motivo de as buscas estarem em nível recorde.[8] O Gemini 3.5 Flash virou, no mesmo dia do lançamento, o modelo padrão do aplicativo Gemini e do AI Mode no mundo todo.[2] Nenhum concorrente consegue colocar um modelo novo diante de tanta gente em um único dia.

Chips e infraestrutura. O Google projeta os próprios chips de IA, os TPUs. Em abril de 2026, apresentou a oitava geração, com dois chips distintos, um para treinar modelos e outro para responder rapidamente a agentes.[9] E a escala do investimento é difícil de acompanhar: no segundo trimestre de 2026, a Alphabet gastou 44,9 bilhões de dólares em equipamentos e instalações, o dobro do mesmo período do ano anterior, e a receita do Google Cloud cresceu 82%, puxada também pela venda de sistemas de TPU.[10]

O concorrente que é cliente. Aqui está o detalhe que mais me impressiona. Em abril de 2026, a Anthropic anunciou um acordo com o Google e a Broadcom para vários gigawatts de capacidade em TPUs, a partir de 2027.[11] Uma das empresas que disputam a fronteira com o Google também roda parte da sua inteligência na infraestrutura do Google. Na corrida do ouro, quem vende as pás ganha mesmo quando outro encontra o ouro.

Ciência e história. A arquitetura Transformer, base de quase todos os grandes modelos de linguagem, foi criada no Google, em 2017.[12] Boa parte da ciência que move esta revolução saiu de lá.

E há o próprio império. No segundo trimestre de 2026, a receita da Busca e de outros serviços cresceu 17%, para 63,3 bilhões de dólares.[10] Quem fala em declínio precisa explicar esse número.

NÃO CONFUNDA

Ficar atrás em um ranking não é ficar fora da corrida. E visitas a sites não são uso total: o levantamento da Similarweb conta só a web, sem os aplicativos de celular nem o Gemini embutido no Android e na Busca, o que tende a subestimar justamente o Google.[13]

3. O risco de proteger o império

Então o Google está bem? Aqui entra a parte desconfortável, a que eu aprendi a reconhecer nas transições que vivi.

O maior risco de uma empresa líder raramente é falta de tecnologia. É ter muito a perder. Cada resposta de IA que resolve a dúvida dentro da página pode significar menos cliques em anúncios e menos tráfego para os sites que alimentam a própria Busca. O Google precisa construir a próxima interface sem destruir, rápido demais, a máquina que paga por ela. Os rivais não têm esse dilema. A OpenAI, que nunca teve um negócio de busca a proteger, começou em 2026 a vender publicidade dentro do ChatGPT, como vimos no capítulo anterior.

E as pessoas estão escolhendo. Em agosto de 2026, segundo a Similarweb, o ChatGPT recebia 55,5% das visitas a sites de IA generativa, e o Gemini, 25,6%. Um ano antes, o Gemini tinha

12,9%: mais que dobrou. Mas, nos três meses anteriores, tinha recuado de 27,8%.[13] O Google cresce, mas ainda não converteu a sua distribuição gigantesca em preferência, pelo menos na web.

Há também o sinal das pessoas. Em junho de 2026, Noam Shazeer, um dos líderes do Gemini e um dos autores do artigo que apresentou o Transformer, anunciou que deixaria o Google para ir para a OpenAI; dias depois, John Jumper, o cientista do AlphaFold premiado com o Nobel, anunciou sua ida para a Anthropic.[14] Em 5 de agosto, a Alphabet anunciou que Demis Hassabis deixaria o comando do Google DeepMind para se tornar presidente do conselho do laboratório e cientista-chefe da Alphabet, com Koray Kavukcuoglu à frente da operação; no mesmo dia, Jeff Dean, Sanjay Ghemawat, Oriol Vinyals e Quoc Le saíram para fundar uma startup própria. Segundo a imprensa, Sundar Pichai tentou mantê-los e respondeu: "Temos talentos incríveis, computação de classe mundial e produtos que levam a IA a mais pessoas do que qualquer outra empresa".[3,15] As duas coisas podem ser verdade ao mesmo tempo. Vamos voltar ao movimento dos talentos no Capítulo 14.

Empresas líderes, com tecnologia excelente, muitas vezes demoram a aceitar uma mudança de regra justamente porque têm muito a proteger.

Eu vivi isso de perto com os celulares, e vamos falar dessa história no Capítulo 16. A lição que levo de lá não é que os líderes caem. É que eles raramente caem por falta de capacidade técnica. Caem quando a interface muda e eles continuam excelentes nas regras anteriores.

Ser frontier não é o mesmo que ser preferido.
Ter o mapa não é o mesmo que ter o cliente.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: é como o time mais rico do campeonato, com o maior estádio e a maior torcida, que perdeu alguns craques e ficou um tempo sem ganhar o clássico. Ninguém sério diz que ele vai ser rebaixado. Mas ninguém garante o título.

Para quem tem 65 anos: é como a grande loja de departamentos do centro quando os shoppings chegaram. Tinha os melhores fornecedores, o melhor ponto e os clientes mais fiéis. A pergunta nunca foi se ela sabia vender; foi se ia aceitar vender de outro jeito.

O QUE AINDA NÃO SABEMOS

Quando o Gemini 4 Argon chegará ao público, e se manterá a liderança em preferência com mais votos. Se a saída de líderes do Gemini vai atrasar os próximos modelos ou apenas renovar a equipe. Se a IA na Busca vai ampliar ou corroer o negócio de anúncios. E se a vantagem de custo e distribuição pesará mais, a longo prazo, do que alguns pontos a mais em um índice.

O que mudou / O que significa / O que fazer agora

O que mudou: em 2026, o Google atrasou o seu modelo mais avançado, perdeu líderes importantes e viu concorrentes lançarem novas gerações antes dele. Ao mesmo tempo, continuou crescendo em receita, nuvem, chips e distribuição, e anunciou um novo modelo de fronteira.

O que significa: a pergunta "o Google está ficando para trás?" não tem uma resposta única. Em capacidade de topo, está um passo atrás neste momento. Em velocidade, custo,

multimodalidade, infraestrutura e alcance, continua entre os mais fortes do mundo. O desafio é transformar essa força em preferência, sem sacrificar o império que a financia.

O que fazer agora: se você usa o Google, teste o Gemini na tarefa que importa para você, e não pela manchete da semana. E acompanhe três sinais: quando o melhor modelo chega ao público, para onde vão os talentos e como evolui a Busca. Google não é Nokia, mas nenhum império recebe imunidade contra uma mudança de interface.

AÇÃO EM DEZ MINUTOS

Faça a mesma pergunta de trabalho em três lugares: na Busca do Google com o AI Mode, no aplicativo Gemini e no assistente que você mais usa. Anote no seu Diário de Fronteira, com a data, qual resposta foi mais útil e qual você usaria de novo. Depois, responda a si mesmo: você escolheu pela qualidade ou pelo hábito?

E se até um gigante americano pode parecer atrás por alguns meses, imagine a velocidade do outro lado do mundo. Muita gente ainda resume a corrida chinesa em um único nome. É o erro que o próximo capítulo vai desfazer: a China não é apenas DeepSeek.

Notas

1. BBC News. "Google's Bard AI bot mistake wipes $100bn off shares." 8 fev. 2023. https://www.bbc.co.uk/news/business-64576225

2. Kavukcuoglu, Koray et al. "Gemini 3.5: frontier intelligence with action." Google, 19 maio 2026. https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-3-5/

3. Whitwam, Ryan. "Google's AI shake-up: DeepMind's Hassabis steps aside, senior scientists depart." Ars Technica, 5 ago. 2026.

https://arstechnica.com/gadgets/2026/08/googles-ai-shakeup-deepminds-hassabis-steps-aside-senior-scientists-depart/

4. Sinha, Debjit. "Google-parent Alphabet unveils Gemini 4 flagship AI model: Check features, release date, Gemini 3.5 Pro scrapped." The Economic Times, com informações da Reuters, 1 out. 2026. https://economictimes.indiatimes.com/news/international/us/google-parent-alphabet-unveils-gemini-4-flagship-ai-model-check-features-release-date-gemini-3-5-pro-scrapped/articleshow/134605546.cms

5. Artificial Analysis. "LLM Leaderboard" (Intelligence Index v4.3.2). Consultado em 6 out. 2026. https://artificialanalysis.ai/leaderboards/models

6. Artificial Analysis. "Gemini 3.6 Flash (high) – Intelligence, Performance & Price Analysis." Consultado em 6 out. 2026. https://artificialanalysis.ai/models/gemini-3-6-flash; Google DeepMind, "Gemini 3.6 Flash – Model Card." 21 jul. 2026. https://deepmind.google/models/model-cards/gemini-3-6-flash/

7. LMArena. "Text Arena – Overall." Atualização de 2 out. 2026, consultada em 6 out. 2026. https://arena.ai/leaderboard/text/overall

8. Mohan, Shivani. "How AI Mode is changing the way people search in the U.S." Google, 19 maio 2026. https://blog.google/products-and-platforms/products/search/ai-mode-us-insights/

9. Google. "We're launching two specialized TPUs for the agentic era." 22 abr. 2026. https://blog.google/innovation-and-ai/infrastructure-and-cloud/google-cloud/tpus-8t-8i-cloud-next/

10. Alphabet. "Alphabet Announces Second Quarter 2026 Results." 22 jul. 2026 (Exhibit 99.1, SEC). https://www.sec.gov/Archives/edgar/data/1652044/000165204426000066/googexhibit991q22026.htm

11. Anthropic. "Anthropic expands partnership with Google and Broadcom for multiple gigawatts of next-generation compute." 6 abr. 2026. https://www.anthropic.com/news/google-broadcom-partnership-compute

12. Google DeepMind. "About Google DeepMind." Consultado em 6 out. 2026. https://deepmind.google/about/

13. Similarweb. "Gen AI website traffic share update – August 2026." 7 set. 2026. https://www.linkedin.com/posts/similarweb_gen-ai-website-traffic-share-update-august-activity-7502646378943401985-V6qg; The Decoder, "ChatGPT claws back web traffic share to 55.5 percent as Gemini's brief comeback fades." 7 set. 2026. https://the-decoder.com/chatgpt-claws-back-web-traffic-share-to-55-5-percent-as-geminis-brief-comeback-fades/

14. Reuters. "US scientist John Jumper to leave Google DeepMind for Anthropic." 19 jun. 2026. https://www.reuters.com/technology/us-scientist-john-jumper-leave-google-deepmind-anthropic-2026-06-19/

15. Reuters. "Google shakes up AI leadership as DeepMind chief shifts role." 5 ago. 2026. https://www.reuters.com/business/google-shakes-up-ai-leadership-deepmind-chief-shifts-role-2026-08-05/

CAPÍTULO 11

A China não é apenas DeepSeek

Tese do capítulo: a corrida chinesa combina modelos próximos da fronteira, abertura de pesos, eficiência, competição interna e implantação industrial em enorme escala.

Terminei o capítulo anterior prometendo desfazer um erro. Para isso, voltei à noite em que este livro começou: eu diante da televisão, ouvindo que a China tinha simulado milhares de AGIs.

No Capítulo 1, já vimos que não eram AGIs, e sim mais de um bilhão de agentes simplificados. Mas, ao reler o artigo para escrever este capítulo, percebi um segundo detalhe que a manchete escondia. Os autores eram de instituições chinesas, como a Academia Zhongguancun, a Universidade de Ciência e Tecnologia da China e a Universidade Tsinghua. Só que o "cérebro" usado em boa parte dos experimentos era o Gemini 2.0 Flash, do Google, comparado com um modelo da OpenAI. Um modelo pequeno da família Qwen, da Alibaba, aparecia como substituto mais barato, e um experimento usava o DeepSeek.[1]

A manchete dizia "a China". O artigo mostrava pesquisadores chineses combinando modelos americanos e chineses no mesmo laboratório. Era um mapa muito mais interessante, e muito mais preciso.

Para muita gente, porém, toda a inteligência artificial chinesa ainda cabe num único nome: DeepSeek.

Entendo por que esse nome ficou. Em 27 de janeiro de 2025, a reação do mercado ao DeepSeek fez a Nvidia perder cerca de 589 bilhões de dólares em valor de mercado em um só dia, a maior perda já registrada por uma empresa americana.[2] Um choque desse tamanho grava um nome na memória. O problema é quando o nome vira o mapa inteiro.

1. DeepSeek, Qwen, Kimi e GLM: quatro apostas diferentes

No Capítulo 8, apresentei a ficha de cada um desses laboratórios. Aqui quero mostrar outra coisa: a lógica de cada um. Por baixo, os quatro usam a mesma família de tecnologia, os modelos Transformer, quase sempre em arquiteturas de "mistura de especialistas", que ativam só uma parte do modelo a cada resposta. O que os diferencia é a estratégia.

DeepSeek: a eficiência como arma. A fama do DeepSeek começou antes do choque de 2025. Em maio de 2024, o preço baixo do seu modelo V2 ajudou a disparar uma guerra de preços na China: ByteDance, Alibaba e Baidu cortaram valores, em alguns casos em mais de 90%, ou ofereceram versões gratuitas.[3] Fazer mais com menos virou a assinatura da casa. E a ambição cresceu: em setembro de 2026, segundo relatos da imprensa, o fundador Liang Wenfeng disse a investidores que a empresa treinava um modelo de 2 trilhões de parâmetros e pretendia usar mais chips chineses, incluindo os da Huawei, no treinamento.[4]

Qwen: a cobertura total. A Alibaba fez o oposto de apostar em um único modelo. Publica versões de quase todos os tamanhos, das pequenas, que rodam em um computador comum, ao

Qwen3.8-Max, com 2,4 trilhões de parâmetros. Resultado: segundo o relatório de agosto de 2026 do Hugging Face, maior repositório de modelos abertos do mundo, havia 151.448 modelos derivados do Qwen na plataforma, 2,6 vezes toda a presença da Meta.[5] A Alibaba diz ter passado de 3 bilhões de downloads; o Hugging Face contou cerca de 2 bilhões na própria plataforma em 2026. É a diferença entre o número da empresa e o número medido por terceiros.[5]

Kimi: só a fronteira. A Moonshot quase não publica modelos pequenos. Aposta tudo no topo, como o Kimi K3, com cerca de 2,8 trilhões de parâmetros.[5] No ranking do LMArena de 2 de outubro de 2026, considerando o melhor modelo de cada laboratório, a Moonshot aparecia em 4.º lugar, à frente da OpenAI.[6] Mas os downloads dos seus modelos em 2026, 37 milhões, eram cerca de 55 vezes menores que os do Qwen.[5] E o aplicativo Kimi tinha cerca de 7 milhões de usuários mensais em junho, contra 382 milhões do Doubao.[7] Brilhar na fronteira não é o mesmo que chegar às pessoas.

GLM: o caminho institucional. A Zhipu, hoje Z.ai, entrou em janeiro de 2025 na lista de restrições comerciais do governo americano, que exige licença para a venda de tecnologia dos Estados Unidos à empresa.[8] Um ano depois, em 8 de janeiro de 2026, abriu capital em Hong Kong. No dia seguinte, foi a vez da MiniMax, outro laboratório chinês. As duas chegaram à bolsa antes da OpenAI.[9]

E a lista não acaba aí. Xiaomi, Ant Group e Meituan, empresas que um ano antes quase ninguém associava a modelos abertos, passaram de 1 trilhão de parâmetros em 2026. Segundo o Hugging Face, em quase todos os meses do ano, o maior modelo aberto chinês foi maior do que qualquer modelo aberto lançado por um laboratório americano.[5]

FOTOGRAFIA DA FRONTEIRA — A CHINA ALÉM DE UM NOME (6 DE OUTUBRO DE 2026)

Preferência (LMArena, 2 out.): Kimi K3, 16.º lugar; Qwen3.8-Max, 23.º; MiMo-V2.6-Pro (Xiaomi), 26.º; GLM-5.3, 27.º; DeepSeek-V4.1-Flash, 38.º; ERNIE 5.1, 48.º; Seed 2.0 Pro (ByteDance), 66.º; Hy3 (Tencent), 67.º.

Usuários de aplicativos (QuestMobile, junho de 2026): Doubao, 382 milhões por mês; Qwen, 167 milhões; DeepSeek, 130 milhões; Yuanbao (Tencent), 50 milhões; Kimi, 7 milhões.

Fatos vivos: confira o Atlas Vivo da IA.

2. ByteDance, Tencent e Baidu: a IA que já tem para onde ir

Se os quatro laboratórios disputam a fronteira, os três gigantes da internet chinesa disputam outra coisa: o lugar onde a IA encontra a vida das pessoas. E eles já têm esse lugar.

ByteDance: conteúdo. A dona do TikTok entende de atenção como poucas empresas no mundo. Em 16 de fevereiro de 2026, durante a Gala do Festival da Primavera da TV estatal chinesa, um dos programas mais assistidos do país, o Doubao passou de 100 milhões de usuários diários. Segundo a ByteDance, recebeu mais de 1,9 bilhão de perguntas durante o programa.[10] Por trás do aplicativo está a Volcano Engine, a nuvem da empresa. Em junho de 2026, seu presidente informou que o modelo Doubao processava mais de 180 trilhões de tokens por dia, dez vezes mais que um ano antes.[11]

Tencent: comunicação e jogos. A Tencent tem o Weixin, que fora da China se chama WeChat, com 1,44 bilhão de usuários

mensais. Em agosto de 2026, informou que começou a testar, em pequena escala, o Xiaowei, um agente de IA dentro do aplicativo. Disse também que o seu modelo Hy3 estava, desde julho, entre os três mais usados do mundo na OpenRouter, plataforma que reúne modelos de vários laboratórios. São informações da própria empresa. Os números financeiros mostram a aposta: a receita de jogos no mercado chinês cresceu 17% no trimestre, e o investimento em infraestrutura saltou 176%, com a empresa comprando muito mais capacidade de computação para IA.[12]

Baidu: busca e carros. O Baidu vive um dilema parecido com o do Google, que vimos no Capítulo 10. No segundo trimestre de 2026, os negócios ligados à IA já respondiam por metade da receita principal, mas a receita de publicidade on-line caiu 19% em um ano. Enquanto isso, o seu serviço de táxis sem motorista, o Apollo Go, chegava a 28 cidades e acumulava mais de 240 milhões de quilômetros sem ninguém ao volante, inclusive com operação comercial em Dubai.[13] É a IA saindo da tela e indo para a rua.

NÃO CONFUNDA

Usuários de aplicativo não medem a capacidade de um modelo. Os números da QuestMobile são estimativas de um painel, cobrem só aplicativos baixados por iniciativa do usuário, sem os assistentes que já vêm instalados no celular, e se referem à China continental.[7] E "aberto" não quer dizer "sem regras": na China, todo serviço de IA para o público segue regras próprias de conteúdo.

3. Nem imitadora, nem só ameaça

O Ocidente costuma errar sobre a China de dois jeitos opostos.

O primeiro erro é enxergar uma imitadora. O AI Index 2026, da Universidade Stanford, mostra que a China lidera em volume de publicações científicas sobre IA, citações e patentes concedidas. Entre os cem artigos de IA mais citados, 41 eram chineses em 2024, contra 33 em 2021.[14] Os Estados Unidos ainda produzem mais modelos de destaque, 59 contra 35 em 2025, e patentes de maior impacto.[14] Mas, como vimos no Capítulo 6, a distância de desempenho entre os melhores modelos dos dois países praticamente desapareceu. Quem imita não lidera em pesquisa nem constrói os maiores modelos abertos do mundo.

O segundo erro é enxergar apenas uma ameaça. Os riscos são reais, e o Capítulo 8 tratou de privacidade e segurança. Mas boa parte da força chinesa chega ao mundo como ferramenta. Desenvolvedores do mundo inteiro, inclusive americanos, constroem produtos sobre modelos chineses. O próprio Hugging Face registrou que alguns grandes modelos abertos americanos de 2026 foram construídos sobre modelos chineses ou usaram componentes deles.[5] Tratar tudo como ameaça é abrir mão de entender, e de usar com critério, uma parte importante da fronteira.

Ver a China com precisão também exige enxergar os seus limites. Há três que eu acompanho.

Chips. Os controles de exportação americanos restringem o acesso da China aos chips mais avançados da Nvidia. Em setembro de 2026, Eric Xu, um dos presidentes rotativos da Huawei, afirmou que a empresa não consegue produzir chips de IA suficientes nem para a demanda chinesa.[15] E, segundo fontes ouvidas pela imprensa, os principais modelos chineses ainda dependem muito de chips da Nvidia para o treinamento, mesmo com o esforço de autossuficiência.[16]

Regras. Desde 2023, as regras chinesas para serviços de IA generativa oferecidos ao público exigem que eles sigam os "valores socialistas fundamentais" e proíbem conteúdos considerados ameaça à segurança nacional. Desde 1.º de setembro de 2025, conteúdos gerados por IA precisam ser identificados.[17] Para quem usa esses modelos, isso tem consequência prática: as respostas sobre certos temas refletem essas regras.

Escala. Aqui o limite vira vantagem. A China não pensa a IA apenas como chatbot, e sim como parte da indústria. Em 2024, instalou cerca de 295 mil robôs industriais, 54% de todos os instalados no mundo, e passou de 2 milhões em operação.[18] Em agosto de 2025, o governo lançou o plano "IA+", com a meta de que os "terminais inteligentes de nova geração", como celulares e computadores com IA, e os agentes de IA passem de 70% de penetração até 2027 e de 90% até 2030.[19]

Quem vê só imitação subestima a China. Quem vê só ameaça deixa de entendê-la. Nos dois casos, o mapa sai errado.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: é como achar que um país só tem um jogador bom porque ele saiu em todas as manchetes. Quando você olha o elenco, descobre um time inteiro, com atacantes, goleiros e categorias de base.

Para quem tem 65 anos: é como quem, nos anos 1980, achava que os produtos japoneses eram só cópias baratas. Poucos anos depois, eram eles que definiam o padrão de qualidade.

O QUE AINDA NÃO SABEMOS

Se os chips chineses vão alcançar os da Nvidia a tempo de sustentar modelos ainda maiores. Se os laboratórios chineses vão manter os pesos abertos quando precisarem de mais receita: alguns já começaram a incluir restrições comerciais nas licenças dos maiores modelos.[5] Se a guerra de preços deixará empresas saudáveis ou só sobreviventes. E como as regras de conteúdo vão afetar o uso desses modelos fora da China.

O que mudou / O que significa / O que fazer agora

O que mudou: a China deixou de ser um nome e passou a ser um ecossistema. Há laboratórios que disputam a fronteira, gigantes da internet que levam a IA a centenas de milhões de pessoas, empresas que abriram capital e uma estratégia de Estado para espalhar a IA pela indústria.

O que significa: a corrida chinesa não tem um único estilo. Combina eficiência, abertura de pesos, competição interna feroz e implantação industrial em escala enorme, e convive com limites reais de chips, de regras e de dados. Nem cópia, nem monstro: um competidor completo, com outra lógica.

O que fazer agora: quando ouvir uma notícia sobre IA chinesa, pergunte qual empresa, qual modelo, qual dado e qual data. Se sua empresa usa modelos abertos, verifique quantos deles têm origem chinesa e em que condições são usados. E inclua a China em qualquer análise de fronteira que você fizer. Ignorar a China produz um mapa falso da fronteira.

AÇÃO EM DEZ MINUTOS

Escreva no seu Diário de Fronteira os nomes de empresas chinesas de IA que você conhece, sem pesquisar. Depois, compare com os laboratórios e as empresas deste capítulo. Anote quais faltavam e qual delas mais surpreendeu você. Por fim, escolha uma para acompanhar nos próximos três meses.

E se cada empresa chinesa joga com uma lógica diferente, o mesmo vale para os países. Estados Unidos, China, Europa e outros não disputam apenas o chatbot mais inteligente. Disputam chips, energia, padrões, indústria e soberania. É o tabuleiro do próximo capítulo: quatro estratégias nacionais para a mesma revolução.

Notas

1. Guan, Haoxiang et al. "Modeling Earth-Scale Human-Like Societies with One Billion Agents." arXiv:2506.12078, jun. 2025 (versão revisada consultada em 6 out. 2026). https://arxiv.org/abs/2506.12078

2. Reuters. "DeepSeek sparks AI stock selloff; Nvidia posts record market-cap loss." 27 jan. 2025. https://www.reuters.com/technology/chinas-deepseek-sets-off-ai-market-rout-2025-01-27/; Bloomberg, "Nvidia's $589 Billion DeepSeek Plunge Is Largest in Market History." 27 jan. 2025. https://www.bloomberg.com/news/articles/2025-01-27/asml-sinks-as-china-ai-startup-triggers-panic-in-tech-stocks

3. Reuters. "Chinese tech giants slash prices of language models used to power AI chatbots." 21 maio 2024. https://www.reuters.com/technology/chinese-tech-giants-slash-prices-language-models-used-power-ai-chatbots-2024-05-21/; Fortune, "China's AI race heats up as ByteDance, Baidu, Alibaba trigger LLM price war." 22 maio 2024. https://fortune.com/asia/2024/05/22/china-ai-llm-price-war-discounts-bytedance-alibaba-baidu-iflytek/

4. TrendForce. "DeepSeek Reportedly Expands Huawei Chips to AI Training; New Supply Could Arrive in 4Q26." 23 set. 2026 (com base em The Information e Chosun Biz). https://www.trendforce.com/news/2026/09/23/news-deepseek-reportedly-expands-huawei-chips-to-ai-training-new-supply-could-arrive-in-4q26/

5. Yakefu, Adina; Solaiman, Irene. "State of Open Models: Summer 2026 Observations." Hugging Face, 14 ago. 2026. https://huggingface.co/blog/state-of-open-models-summer-2026; sobre a cifra da Alibaba: The Next Web, "Qwen is the world's most downloaded open model, by a smaller margin than Alibaba says." Ago. 2026. https://thenextweb.com/news/alibaba-qwen-downloads-hugging-face-open-models

6. LMArena. "Text Arena – Overall" (classificação por laboratório). Atualização de 2 out. 2026, consultada em 6 out. 2026. https://arena.ai/leaderboard/text/overall

7. QuestMobile, ranking de aplicativos nativos de IA do primeiro semestre de 2026 (dados de junho de 2026), conforme Kuai Keji/iFeng, "中国 AI 原生 APP 用户量 TOP10 公布：豆包 3.82 亿月活断层第一." 5 ago. 2026. https://i.ifeng.com/c/8vLUETsPJuY

8. U.S. Department of Commerce, Bureau of Industry and Security. "Addition of Entities to and Revision of Entry on the Entity List." Federal Register, 16 jan. 2025. https://www.federalregister.gov/documents/2025/01/16/2025-00704/addition-of-entities-to-and-revision-of-entry-on-the-entity-list

9. Butts, Dylan. "MiniMax doubles in Hong Kong debut, marking yet another Chinese AI listing." CNBC, 9 jan. 2026. https://www.cnbc.com/2026/01/09/minimax-hong-kong-ipo-ai-tigers-zhipu.html; Z.ai, "China's AGI Pioneer and Leader Z.ai Listed on Hong Kong Stock Exchange." PR Newswire, jan. 2026. https://www.prnewswire.com/apac/news-releases/chinas-agi-pioneer-and-leader-zai-listed-on-hong-kong-stock-exchange-302656269.html

10. Reuters. "ByteDance's AI chatbot Doubao wins China holiday battle with 100 million users." 25 fev. 2026. https://www.reuters.com/world/asia-

pacific/bytedances-ai-chatbot-doubao-wins-china-holiday-battle-with-100-million-users-2026-02-25/

11. Jiemian Global. "Doubao model's daily token usage hits 180 trillion." 23 jun. 2026. https://en.jiemian.com/article/14629930.html

12. Tencent. "Tencent Announces 2026 Second Quarter Results." 12 ago. 2026. https://www.prnewswire.com/apac/news-releases/tencent-announces-2026-second-quarter-results-302849608.html

13. Baidu. "Baidu Announces Second Quarter 2026 Results." 18 ago. 2026 (Exhibit 99.1, SEC). https://www.sec.gov/Archives/edgar/data/1329099/000119312526355431/d112854dex991.htm

14. Stanford HAI. "The 2026 AI Index Report", cap. 1, "Research and Development." 2026. https://hai.stanford.edu/ai-index/2026-ai-index-report/research-and-development

15. Pan, Che; Hall, Casey; Baptista, Eduardo. "China's Huawei says AI chip demand outstrips supply as it steps up Nvidia challenge." Reuters, 17 set. 2026 (reproduzido por KFGO). https://kfgo.com/2026/09/16/chinas-huawei-to-launch-two-new-ai-chips-in-2027/

16. Gupta, Sahil Mohan. "Top Chinese AI Models Trained On Nvidia Despite Self Sufficiency Push." Gadgets Now, 10 ago. 2026. https://gadgetsnow.indiatimes.com/tech-news/top-chinese-ai-models-trained-on-nvidia-despite-self-sufficiency-push/articleshow/133117788.cms

17. Administração do Ciberespaço da China (CAC) e outros seis órgãos. "生成式人工智能服务管理暂行办法" (Medidas provisórias para a gestão de serviços de IA generativa), art. 4.º; publicadas em 10 jul. 2023, em vigor desde 15 ago. 2023. Texto em: https://press.nppa.gov.cn/zcfg/202506/t20250625_918911.html; CAC, MIIT, Ministério da Segurança Pública e NRTA. "人工智能生成合成内容标识办法" (Medidas para a identificação de conteúdo gerado ou sintetizado por IA). 14 mar. 2025, em vigor desde 1.º set. 2025. https://www.cac.gov.cn/2025-03/14/c_1743654684782215.htm

18. International Federation of Robotics. "China Tops World Record of 2 Million Factory Robots." 25 set. 2025.

https://ifr.org/downloads/press_docs/2025-09-25-IFR_press_release_China_in_English.pdf

19. Conselho de Estado da República Popular da China. "Opiniões sobre o aprofundamento da iniciativa 'Inteligência Artificial+'" (Guofa [2025] n.º 11). 21 ago. 2025, publicado em 26 ago. 2025; resumo oficial em inglês: "China issues guideline to accelerate 'AI Plus' integration across key sectors." 27 ago. 2025. https://english.www.gov.cn/policies/latestreleases/202508/27/content_WS68ae7976c6d0868f4e8f51a0.html; tradução: CSET, Georgetown, 24 set. 2025. https://cset.georgetown.edu/publication/china-ai-plus-opinions-2025/

CAPÍTULO 12

Quatro estratégias nacionais para a mesma revolução

Tese do capítulo: países não disputam apenas o chatbot mais inteligente; disputam infraestrutura, padrões, indústria, segurança, soberania e influência.

Terminei o capítulo anterior com uma constatação: se cada empresa chinesa joga com uma lógica diferente, o mesmo vale para os países. Eu sinto isso na pele há muitos anos, porque vivo entre dois deles.

Quem trabalha entre o Brasil e os Estados Unidos aprende cedo que a mesma tecnologia chega a lugares diferentes em velocidades diferentes, com preços diferentes e regras diferentes. Um produto lançado nos Estados Unidos em uma semana pode levar meses para chegar ao Brasil, e chegar mais caro, com outra embalagem e outro contrato. Não é falta de talento. É diferença de recursos, de escala e de urgência.

Com a inteligência artificial, essa diferença ficou maior e mais estratégica. Os países não estão apenas escolhendo qual chatbot usar. Estão decidindo onde ficarão os data centers, quem fabricará os chips, quais regras valerão, quem formará os talentos e de quem o resto do mundo vai depender. Neste capítulo, olho para quatro estratégias: Estados Unidos, China, União Europeia e Reino Unido. E, no fim, para o lugar onde muitos leitores deste livro estão: o Brasil.

1. Estados Unidos: a fronteira privada

A estratégia americana cabe numa ideia do próprio governo: criar as condições para que a inovação liderada pelo setor privado floresça, e tirar obstáculos do caminho. Em julho de 2025, a Casa Branca publicou o Plano de Ação de IA dos Estados Unidos, com mais de 90 medidas organizadas em três pilares: acelerar a inovação, construir infraestrutura e liderar na diplomacia e na segurança internacionais.[1] O texto fala em remover "burocracia e regulação onerosa" e em vencer a corrida.

Quem corre, porém, é o capital privado. Segundo o AI Index 2026, da Universidade Stanford, o investimento privado em IA nos Estados Unidos chegou a 285,9 bilhões de dólares em 2025, mais de 23 vezes o da China.[2] E esse número nem inclui o maior gasto de todos: o das grandes plataformas em infraestrutura. Depois dos resultados de julho de 2026, Amazon, Google, Microsoft e Meta projetavam juntas mais de 735 bilhões de dólares em investimentos no ano, grande parte em data centers e chips.[3] Os Estados Unidos já tinham 5.427 data centers, mais de dez vezes qualquer outro país.[4]

A segunda peça é o controle dos chips. O governo americano usa o acesso aos semicondutores mais avançados como instrumento de política externa. Em janeiro de 2026, por exemplo, passou a analisar caso a caso a exportação para a China de chips como o H200, da Nvidia, que antes tinham, em regra, o pedido negado. Mas impôs exigências e limites rígidos.[5] É uma torneira que se abre e se fecha conforme a estratégia.

A terceira peça é exportar o pacote inteiro. O plano americano quer que outros países adotem não só modelos americanos, mas toda a pilha: chips, nuvem, modelos e aplicações. Em julho de 2026, o Departamento de Comércio informou ter recebido 78

propostas de consórcios de empresas para exportar esses pacotes completos, com apoio do governo.[6]

E há uma fragilidade que me chamou atenção. O mesmo AI Index registrou que o número de pesquisadores e desenvolvedores de IA que se mudam para os Estados Unidos caiu 89% desde 2017, e 80% apenas no último ano.[2] Durante décadas, a maior vantagem americana foi atrair os melhores cérebros do mundo. Se esse ímã enfraquecer, nenhum data center compensa.

2. China: a coordenação nacional

No Capítulo 11, vimos o ecossistema chinês por dentro. Aqui interessa a lógica do Estado. Se a estratégia americana é deixar o mercado correr, a chinesa é coordenar a corrida.

O plano "IA+", do Conselho de Estado (Guofa [2025] n.º 11), definiu metas de adoção para 2027 e 2030, como vimos.[7] Em março de 2026, o 15.º Plano Quinquenal, que orienta a economia chinesa de 2026 a 2030, colocou essa iniciativa no centro e combinou IA com autossuficiência tecnológica em áreas como chips e software.[8] A ideia é levar a IA a cada fábrica, hospital e repartição, não apenas ao celular.

A China também aposta na educação desde cedo. Em maio de 2025, o Ministério da Educação publicou diretrizes para o ensino de IA nas escolas, com regras que proíbem alunos dos primeiros anos do ensino fundamental de usar sozinhos ferramentas que geram conteúdo livremente.[9] Em Pequim, desde o segundo semestre de 2025, todas as escolas de ensino fundamental e médio precisam oferecer pelo menos oito aulas de IA por ano letivo.[10]

E há um detalhe que os números escondem. O investimento privado chinês parece pequeno perto do americano, mas o próprio

AI Index alerta que esse dado provavelmente subestima o gasto total da China, porque fundos de orientação do governo aplicaram, segundo estimativas, cerca de 184 bilhões de dólares em empresas de IA entre 2000 e 2023.[11] Comparar só o capital privado dos dois países é medir uma corrida olhando apenas para um dos corredores.

NÃO CONFUNDA

Plano não é resultado. Uma meta oficial de adoção diz o que um governo quer, não o que já aconteceu. E investimento privado não é investimento total: em países onde o Estado investe diretamente, os rankings de capital privado contam só uma parte da história.

3. Europa e Reino Unido: confiança, soberania e capacidade

União Europeia: primeiro as regras, agora a capacidade. Em 2024, a Europa aprovou uma lei abrangente de IA, o AI Act, com aplicação gradual. As proibições de usos considerados inaceitáveis valem desde fevereiro de 2025, e as obrigações para os modelos de IA de uso geral, desde agosto de 2025. Em 2 de agosto de 2026, a maior parte das regras passou a valer, e a fiscalização começou.[12]

Mas houve uma mudança importante no caminho. Em julho de 2026, a União Europeia adotou uma "simplificação" do AI Act, o chamado Digital Omnibus, que adiou as regras para sistemas de alto risco, como os usados em seleção de pessoal ou concessão de crédito. Elas passaram a valer só em dezembro de 2027, e, no caso de IA embutida em produtos regulados, em agosto de 2028.[13] A

Europa continua apostando na confiança, mas começou a ajustar o ritmo para não sufocar quem constrói.

Ao mesmo tempo, a Europa passou a investir em músculo. Em fevereiro de 2025, a Comissão Europeia lançou o InvestAI, iniciativa para mobilizar 200 bilhões de euros em IA.[14] Em 30 de julho de 2026, abriu a concorrência para até sete "gigafábricas de IA", enormes centros de computação com até 10 bilhões de euros de dinheiro público da UE e dos países-membros e a expectativa de atrair pelo menos 20 bilhões de euros do setor privado. A seleção está prevista para o início de 2027.[15] No Capítulo 8, vimos a Mistral, a principal construtora europeia de modelos. As gigafábricas são a tentativa de dar a empresas como ela uma infraestrutura própria, dentro das regras europeias.

Reino Unido: o laboratório de segurança que quer crescer. Fora da União Europeia, o Reino Unido escolheu um caminho próprio. Em janeiro de 2025, o governo adotou integralmente as 50 recomendações do Plano de Ação de Oportunidades em IA, com metas como multiplicar por 20 a capacidade de computação pública até 2030 e criar "Zonas de Crescimento de IA", com licenças e conexão elétrica mais rápidas para data centers.[16] Um ano depois, informou ter cumprido 38 das 50 ações e designado cinco zonas, com investimentos anunciados de 28,2 bilhões de libras.[17]

O país também transformou a segurança em diferencial. Seu instituto público de avaliação de modelos, criado como AI Safety Institute, passou a se chamar AI Security Institute em fevereiro de 2025, com foco em riscos à segurança nacional.[18] E, em setembro de 2025, assinou com os Estados Unidos um acordo de tecnologia em que empresas americanas anunciaram 31 bilhões de libras de investimento no país.[19] É uma estratégia de ponte: próxima dos Estados Unidos, com voz própria em segurança.

FOTOGRAFIA DA FRONTEIRA — QUATRO ESTRATÉGIAS (6 DE OUTUBRO DE 2026)

Estados Unidos: Plano de Ação de IA (jul. 2025); US$ 285,9 bilhões de investimento privado em 2025; grandes plataformas projetando mais de US$ 735 bilhões em investimentos em 2026.

China: plano "IA+" (Guofa [2025] n.º 11) e 15.º Plano Quinquenal (mar. 2026); IA obrigatória nas escolas de Pequim desde 2025.

União Europeia: AI Act aplicado desde 2 ago. 2026, com alto risco adiado para dez. 2027 e ago. 2028; concorrência das gigafábricas aberta até 12 nov. 2026.

Reino Unido: 38 de 50 ações do plano cumpridas (jan. 2026); cinco Zonas de Crescimento de IA.

Fatos vivos: confira o Atlas Vivo da IA.

E o Brasil? Seria estranho escrever este capítulo sem olhar para casa. Em julho de 2024, o governo brasileiro lançou o Plano Brasileiro de Inteligência Artificial, com previsão de 23 bilhões de reais até 2028, incluindo um supercomputador e aplicações em saúde e serviços públicos.[20] O marco legal da IA, o Projeto de Lei 2338/2023, foi aprovado pelo Senado em dezembro de 2024 e, em outubro de 2026, ainda aguardava parecer em uma comissão especial da Câmara dos Deputados.[21]

Para ter uma ideia da escala, os 23 bilhões de reais do plano brasileiro, distribuídos em quatro anos, são uma fração do que uma única grande plataforma americana gasta em infraestrutura em um trimestre. Isso não torna o plano inútil. Torna a escolha mais importante. Um país com menos recursos não pode tentar fazer tudo. Precisa decidir em que camada quer ser forte, e de quem aceita depender nas outras.

Quem constrói a infraestrutura escreve, em silêncio, parte das regras para todos os outros.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: é como quatro escolas montando seus times para o mesmo campeonato. Uma contrata os melhores jogadores com patrocínio milionário. Outra treina todos os alunos desde pequenos com um plano rígido. Uma terceira escreve o regulamento. A quarta organiza a arbitragem. Todas querem ganhar, cada uma do seu jeito.

Para quem tem 65 anos: é como a corrida das estradas de ferro e da energia elétrica no século passado. Quem construiu os trilhos e as usinas definiu por onde as cidades iam crescer. Os outros tiveram de se adaptar ao traçado.

O QUE AINDA NÃO SABEMOS

Se os investimentos gigantes em data centers vão gerar retorno à altura ou se haverá excesso de capacidade. Se a queda na atração de talentos vai enfraquecer os Estados Unidos. Se o adiamento europeu vai atrair empresas ou só adiar o problema. Se as gigafábricas sairão do papel no prazo. E qual espaço sobrará para países como o Brasil entre as grandes estratégias.

O que mudou / O que significa / O que fazer agora

O que mudou: entre 2025 e 2026, a IA virou política de Estado nas maiores economias do mundo. Os Estados Unidos apostaram no capital privado, nos chips e na exportação do pacote americano. A China coordenou indústria, educação e autonomia tecnológica. A União Europeia começou a aplicar sua lei e a investir em

computação própria. O Reino Unido combinou crescimento com segurança.

O que significa: a corrida da IA não é só uma disputa entre empresas. É também uma disputa entre modelos de sociedade, com escolhas diferentes sobre mercado, Estado, regras e confiança. As decisões tomadas agora sobre energia, chips, data centers e padrões vão definir, por muitos anos, quem depende de quem.

O que fazer agora: se você lidera uma empresa, descubra em que país estão os dados, os modelos e a nuvem que você usa, e sob quais regras eles operam. Se atua no Brasil, acompanhe o marco legal na Câmara e o plano nacional, porque eles vão afetar o seu negócio. E lembre que nenhuma tecnologia chega neutra. Quem define a infraestrutura da inteligência influencia como o resto do mundo aprende, trabalha e decide.

AÇÃO EM DEZ MINUTOS

Liste no seu Diário de Fronteira as três ferramentas de IA que você mais usa. Para cada uma, anote a empresa, o país de origem e onde os seus dados ficam guardados, se a empresa informar. Depois, escreva uma frase: se uma dessas ferramentas deixasse de estar disponível no seu país amanhã, o que você faria?

E se a corrida entre países mostra estratégias diferentes, há um empresário que resolveu juntar, em um só grupo, foguetes, satélites, uma rede social e um laboratório de IA. Para entender o que ele está construindo, não basta olhar para mais um chatbot. É preciso olhar para a pilha inteira: Musk, X, Grok e SpaceX.

Notas

1. The White House. "White House Unveils America's AI Action Plan." 23 jul. 2025. https://www.whitehouse.gov/releases/2025/07/white-house-unveils-americas-ai-action-plan/; "Winning the Race: America's AI Action Plan." Jul. 2025. https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf

2. Stanford HAI. "The 2026 AI Index Report" (principais conclusões: investimento e talentos). 2026. https://hai.stanford.edu/ai-index/2026-ai-index-report

3. Fitzgerald, Charles. "Follow the CAPEX: Q2 2026 Scoreboard." Platformonomics, 31 jul. 2026 (compilação das projeções divulgadas pelas empresas). https://platformonomics.com/2026/07/follow-the-capex-q2-2026-scoreboard/

4. Stanford HAI. "The 2026 AI Index Report", cap. 1, "Research and Development." 2026. https://hai.stanford.edu/ai-index/2026-ai-index-report/research-and-development

5. U.S. Department of Commerce, Bureau of Industry and Security. "Revision to License Review Policy for Advanced Computing Commodities." Federal Register, 15 jan. 2026. https://www.federalregister.gov/documents/2026/01/15/2026-00789/revision-to-license-review-policy-for-advanced-computing-commodities

6. International Trade Administration. "Commerce Statement on Application Conclusion, Next Steps of the American AI Exports Program." 6 jul. 2026. https://www.trade.gov/press-release/commerce-statement-application-conclusion-next-steps-american-ai-exports-program

7. Conselho de Estado da República Popular da China. "Opiniões sobre o aprofundamento da iniciativa 'Inteligência Artificial+'" (Guofa [2025] n.º 11). 21 ago. 2025, publicado em 26 ago. 2025; resumo oficial em inglês: "China issues guideline to accelerate 'AI Plus' integration across key sectors." 27 ago. 2025. https://english.www.gov.cn/policies/latestreleases/202508/27/content_WS68ae7976c6d0868f4e8f51a0.html

8. Xinhua. "中华人民共和国国民经济和社会发展第十五个五年规划纲要" (Esboço do 15.º Plano Quinquenal). 13 mar. 2026. https://www.news.cn/politics/20260313/085af5de5a4b4268aa7d87d90817df2f/c.html; Xinhua, "China's five-year blueprint set to fast-track tech adoption on factory floor." 15 mar. 2026. https://english.news.cn/20260315/f9bc0f36ebc240299d3b2d47a558da29/c.html

9. Global Times, via People's Daily Online. "China issues guidelines to promote AI education in primary and secondary schools." 13 maio 2025. https://en.people.cn/n3/2025/0513/c90000-20313921.html

10. CCTV. "9 月起 北京中小学生都要开始学习这门课" (A partir de setembro, alunos de Pequim começam a estudar esta matéria). 1 jul. 2025. https://news.cctv.com/2025/07/01/ARTIQOFPKSTI08DRMEzXZa5V250701.shtml

11. Stanford HAI. "The 2026 AI Index Report", cap. "Economy." 2026. https://hai.stanford.edu/ai-index/2026-ai-index-report/economy

12. Comissão Europeia, AI Act Service Desk. "Timeline for the Implementation of the EU AI Act." Consultado em 6 out. 2026. https://ai-act-service-desk.ec.europa.eu/en/ai-act/timeline/timeline-implementation-eu-ai-act

13. Regulamento (UE) 2026/1744 do Parlamento Europeu e do Conselho, de 8 de julho de 2026 (Digital Omnibus on AI). Jornal Oficial, 24 jul. 2026, em vigor desde 27 jul. 2026. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32026R1744

14. Comissão Europeia. "EU launches InvestAI initiative to mobilise €200 billion of investment in artificial intelligence." 11 fev. 2025. https://ec.europa.eu/commission/presscorner/api/files/document/print/en/ip_25_467/IP_25_467_EN.pdf

15. Comissão Europeia. "EU launches AI Gigafactories call to boost Europe's computing capacity and unlock more than €30 billion in investment." 30 jul. 2026. https://digital-strategy.ec.europa.eu/en/news/eu-launches-ai-gigafactories-call-boost-europes-computing-capacity-and-unlock-more-eu30-billion; EuroHPC JU, "The EuroHPC Joint Undertaking launches the AI Gigafactories Call." 30 jul. 2026.

https://www.eurohpc-ju.europa.eu/eurohpc-joint-undertaking-launches-ai-gigafactories-call-2026-07-30_en

16. GOV.UK. "Prime Minister sets out blueprint to turbocharge AI." 13 jan. 2025. https://www.gov.uk/government/news/prime-minister-sets-out-blueprint-to-turbocharge-ai; Clifford, Matt. "AI Opportunities Action Plan" (CP 1241). Jan. 2025. https://assets.publishing.service.gov.uk/media/67851771f0528401055d2329/ai_opportunities_action_plan.pdf

17. Department for Science, Innovation and Technology. "AI Opportunities Action Plan: One Year On." 29 jan. 2026. https://www.gov.uk/government/publications/ai-opportunities-action-plan-one-year-on

18. GOV.UK. "Tackling AI security risks to unleash growth and deliver Plan for Change." 14 fev. 2025. https://www.gov.uk/government/news/tackling-ai-security-risks-to-unleash-growth-and-deliver-plan-for-change

19. GOV.UK. "US-UK pact will boost advances in drug discovery, create tens of thousands of jobs and transform lives." Set. 2025. https://www.gov.uk/government/news/us-uk-pact-will-boost-advances-in-drug-discovery-create-tens-of-thousands-of-jobs-and-transform-lives; The White House, "Memorandum of Understanding ... Regarding the Technology Prosperity Deal." 18 set. 2025. https://www.whitehouse.gov/presidential-actions/2025/09/memorandum-of-understanding-between-the-government-of-the-united-states-of-america-and-the-government-of-the-united-kingdom-of-great-britain-and-northern-ireland-regarding-the-technology-prosperity-de/

20. Valor Econômico. "Supercomputador, cursos e apoio a empresas: conheça o plano de R$ 23 bi para desenvolvimento da IA no Brasil." 30 jul. 2024. https://valor.globo.com/brasil/noticia/2024/07/30/supercomputador-cursos-e-apoio-a-empresas-conheca-o-plano-de-r-23-bi-para-desenvolvimento-da-ia-no-brasil.ghtml; MCTI, "IA para o Bem de Todos: Plano Brasileiro de Inteligência Artificial 2024–2028." https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/plano-brasileiro-de-inteligencia-artificial-pbia-%5Fvf.pdf

21. Agência Senado. "Senado aprova regulamentação da inteligência artificial; texto vai à Câmara." 10 dez. 2024. https://www12.senado.leg.br/noticias/materias/2024/12/10/senado-aprova-regulamentacao-da-inteligencia-artificial-texto-vai-a-camara; Câmara dos Deputados. "PL 2338/2023 – Ficha de tramitação." Consultada em 6 out. 2026. https://www.camara.leg.br/proposicoesWeb/fichadetramitacao/fichadetramitacao/?idProposicao=2487262

CAPÍTULO 13

Musk, X, Grok e SpaceX

Tese do capítulo: a estratégia de Musk precisa ser lida como integração de camadas, não apenas como lançamento de outro chatbot.

Terminei o capítulo anterior olhando para países que disputam chips, energia, regras e influência. Agora quero olhar para um empresário que tenta reunir, dentro de um único grupo, várias dessas peças ao mesmo tempo.

No Capítulo 1, contei que acompanhei Nokia e Motorola em mais de 1.500 pontos de venda. Quem trabalha com distribuição sabe que o melhor aparelho nem sempre ganha sozinho. Ganha quem junta o aparelho, o plano, a loja, o vendedor treinado e o cliente passando na porta. Quem controla a distribuição muitas vezes controla o resultado. Por isso, quando leio as manchetes sobre Elon Musk, não pergunto primeiro se o Grok é o melhor chatbot. Pergunto quantas peças da cadeia ele já tem nas mãos.

Este capítulo separa duas coisas que costumam vir misturadas. A primeira é o que está documentado: compras, números, contratos, processos. A segunda é a aposta: o que Musk diz que vai construir. Uma é fato. A outra ainda é hipótese, e merece ser tratada como tal, sem torcida a favor nem contra.

1. O fato: a SpaceX comprou a xAI, que já incluía o X

A história societária cabe em três passos. Em março de 2025, a xAI, laboratório de IA criado por Musk em 2023, incorporou o X, a antiga Twitter. Em 2 de fevereiro de 2026, a SpaceX concluiu a compra da xAI.[1,2] Segundo a Reuters e a CNBC, a operação, feita com troca de ações, avaliou a SpaceX em US$ 1 trilhão e a xAI em US$ 250 bilhões, a maior fusão já registrada.[1] Ao anunciar o negócio, Musk o descreveu como a criação do motor de inovação "mais ambicioso e verticalmente integrado na Terra (e fora dela)" e apontou como principal motivo a construção de "data centers orbitais".[1]

Antes disso, outra empresa de Musk já tinha entrado na história. Em janeiro de 2026, a Tesla informou ter investido US$ 2 bilhões na xAI. A decisão foi polêmica: meses antes, uma consulta não vinculante aos acionistas da Tesla sobre esse investimento não tinha sido aprovada, porque as abstenções contaram como votos contrários. A Tesla seguiu em frente mesmo assim, argumentando que o acordo ajudaria a levar IA ao mundo físico.[3]

Depois vieram o dinheiro e mais uma compra. Em 11 de junho de 2026, a SpaceX precificou sua oferta pública inicial a US$ 135 por ação, levantando US$ 75 bilhões, a maior estreia da história dos Estados Unidos. As ações começaram a ser negociadas no dia seguinte, e, com o lote adicional, a captação chegou a cerca de US$ 85,7 bilhões. Musk manteve mais de 80% do poder de voto.[2,4] Em agosto, a SpaceX concluiu a compra da Cursor, uma das ferramentas de programação com IA mais usadas por desenvolvedores, por um valor implícito de US$ 60 bilhões.[5]

NÃO CONFUNDA

SpaceX é a empresa-mãe, de capital aberto desde junho de 2026.

xAI é o laboratório de IA, hoje uma subsidiária, que assina seus produtos como *SpaceXAI*.

X é a rede social, que pertence à xAI desde março de 2025.

Grok é a família de modelos e o assistente que você usa no aplicativo, no X ou no carro.

Tesla é uma empresa separada, com outros acionistas, que investiu na xAI e tem acordos com ela.

2. A pilha: distribuição, dados em tempo real, modelo, compute, conectividade e capital

Para entender a estratégia, gosto de empilhar as peças, de baixo para cima. Todos os números a seguir vêm dos documentos da própria empresa e devem ser lidos como tal.

Distribuição. Segundo o prospecto entregue à SEC, o X e o Grok somavam cerca de 550 milhões de usuários ativos mensais em 31 de março de 2026. Desses, cerca de 117 milhões usavam os recursos de IA do Grok.[6] O Grok também aparece em carros da Tesla.[3] É a velha lição do ponto de venda: o produto já está onde o cliente passa.

Dados em tempo real. A empresa afirma que o acesso aos dados em tempo real do X, somado ao seu modelo e à sua infraestrutura, cria "uma vantagem estratégica significativa".[6] É uma alegação da empresa, mas faz sentido entender o argumento: poucas empresas de IA têm uma conversa pública acontecendo, minuto a minuto, dentro de casa.

Modelo. O Grok 4.5 foi lançado em julho de 2026, treinado em parceria com a Cursor, segundo a empresa. O Grok 4.7 chegou em setembro, apresentado como o modelo mais capaz da casa para programação e trabalho intelectual.[7] No Capítulo 6, vimos que ele está entre os bons modelos fechados, mas não no topo do índice da Artificial Analysis.

Compute. Aqui está a peça menos visível e talvez a mais importante. A empresa opera os data centers Colossus e Colossus II, em Memphis, no Tennessee, e em Southaven, no Mississippi.[8] No fim de junho de 2026, declarava 1,4 gigawatt de capacidade nominal, com a expectativa de passar de 2 gigawatts até o fim do ano.[9] E começou a alugar essa capacidade até para concorrentes. Segundo o prospecto, a Anthropic se comprometeu, em maio, a pagar US$ 1,25 bilhão por mês até maio de 2029 por acesso a cerca de 325 mil GPUs da Nvidia, embora o contrato possa ser encerrado por qualquer das partes, com aviso de 90 dias, depois dos três primeiros meses. No fim de maio, Musk disse que o compromisso firme era de um aluguel de seis meses.[8] Só no segundo trimestre, a empresa fechou contratos desse tipo que somam US$ 14,1 bilhões.[9]

Conectividade. A Starlink tinha cerca de 9 mil satélites de banda larga em órbita em março de 2026 e chegou a 12 milhões de assinantes no fim de junho, o dobro de um ano antes.[9,10] É uma rede que já chega a mais de 160 países e territórios, e poucas empresas no mundo têm algo parecido.

Capital. É a Starlink que paga boa parte da conta. Em 2025, o segmento de conectividade faturou US$ 11,4 bilhões e teve US$ 4,4 bilhões de lucro operacional, enquanto o segmento de IA faturou US$ 3,2 bilhões e teve US$ 6,4 bilhões de prejuízo operacional.[10] No segundo trimestre de 2026, a receita de IA saltou para US$ 2,6 bilhões, puxada pelo aluguel de computação. O

segmento ainda teve prejuízo operacional de US$ 1,3 bilhão, embora tenha ficado positivo pelo critério de EBITDA ajustado. Só nesse trimestre, investiu US$ 15,8 bilhões em infraestrutura de IA.[9]

Muita gente tem um bom modelo. Pouca gente tem a loja, o cabo, o satélite e o cofre ao mesmo tempo.

FOTOGRAFIA DA FRONTEIRA — A PILHA DE MUSK (6 DE OUTUBRO DE 2026)

Estrutura: SpaceX (Nasdaq: SPCX) controla a xAI, que controla o X; compra da Cursor concluída em ago. 2026.

Modelo: Grok 4.7 (set. 2026); 46 pontos no índice da Artificial Analysis v4.3.2.

Usuários (empresa): cerca de 550 milhões de ativos mensais no X e no Grok; 117 milhões usando a IA do Grok (31 mar. 2026).

Compute (empresa): 1,4 GW no fim de jun. 2026; meta de mais de 2 GW no fim do ano.

Starlink: 12 milhões de assinantes (30 jun. 2026).

Fatos vivos: confira o Atlas Vivo da IA.

3. A aposta: mundo físico, robótica e infraestrutura espacial

Tudo o que vimos até aqui já existe. O que vem agora é promessa, e a diferença importa.

A aposta mais ousada são os data centers no espaço. No prospecto, a SpaceX diz esperar começar a lançar satélites de

computação de IA "já em 2028", em constelações que poderiam chegar a "potencialmente milhões de satélites", alimentados por energia solar e resfriados pelo próprio ambiente espacial.[11] Em janeiro de 2026, no Fórum Econômico Mundial, Musk foi mais longe: disse que o espaço seria o lugar mais barato para colocar IA "dentro de dois anos, talvez três".[12]

Há quem duvide, com bons argumentos. Um estudo publicado em 2026 no repositório científico arXiv por Slava Turyshev, pesquisador do Jet Propulsion Laboratory, ligado à NASA, concluiu que, para competir com data centers em terra no uso geral, o custo somado de lançar e construir esses satélites teria de ser de 3,4 a 13,5 vezes menor que o preço público atual de um lançamento dedicado do Falcon 9, isso antes de contar comunicação, operação e desgaste. O próprio estudo vê usos iniciais mais viáveis em tarefas que já nascem no espaço.[13] Ou seja: não é impossível, mas a conta ainda não fecha para o uso comum.

A segunda aposta é o mundo físico. A SpaceX descreve a Terafab, uma iniciativa de fabricação de chips com a Tesla e a Intel, com dois tipos de chip: um para os robôs Optimus e os carros da Tesla, outro para a computação em órbita. Por enquanto, existe um acordo geral, e os projetos, prazos e investimentos ainda não foram definidos, segundo o próprio documento.[11] A Tesla, por sua vez, diz que o investimento na xAI serve para levar IA "ao mundo físico em escala".[3]

O que a pilha também carrega. Integração acelera, mas também espalha o risco por todas as camadas. Em julho de 2025, o Grok publicou mensagens antissemitas no X e chegou a se chamar de "MechaHitler"; a xAI pediu desculpas e atribuiu o problema a uma atualização de código.[14] Em janeiro de 2026, a ferramenta de imagens do Grok foi usada para criar imagens sexualizadas de mulheres e menores. O regulador britânico Ofcom

abriu investigação em 12 de janeiro, e a xAI restringiu a edição de imagens dias depois.[15] Em 26 de janeiro, a Comissão Europeia abriu uma investigação formal sob a Lei de Serviços Digitais.[16] O prospecto reconhece investigações em vários países e ações coletivas nos Estados Unidos. Também registra que, em dezembro de 2025, a Comissão já tinha multado o X em 120 milhões de euros por outras infrações, decisão que a empresa contestou na Justiça europeia.[17]

Há ainda o custo físico. Em abril de 2026, a NAACP, organização americana de direitos civis, processou a xAI por operar dezenas de turbinas a gás sem licença ambiental em Southaven para alimentar o Colossus II, perto de casas e escolas, segundo a ação. A empresa sustenta que as turbinas são fontes móveis, dispensadas dessa licença, e o governo americano pediu para intervir a seu favor. No fim de setembro, o caso seguia sem decisão final.[18] Nada disso apaga o que foi construído. Mas mostra que, quando uma só empresa controla muitas camadas, um erro em uma delas pode atingir todas as outras.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: imagine alguém dono ao mesmo tempo do jogo, da plataforma onde todo mundo transmite as partidas, da internet da sua casa e da fábrica de consoles. Pode ser incrível, porque tudo conversa. Mas se essa pessoa errar, ou mudar as regras, o problema chega a tudo de uma vez.

Para quem tem 65 anos: lembre das antigas companhias que eram donas da ferrovia, da mina e da cidade operária ao redor. A integração trazia eficiência e poder. E foi justamente por esse poder que, com o tempo, surgiram regras para equilibrar o jogo.

O QUE AINDA NÃO SABEMOS

Se os data centers orbitais vão sair do papel em 2028 e a que custo. Se o aluguel de computação para concorrentes é uma fonte de receita duradoura ou temporária. Se o Grok vai alcançar o topo dos rankings. Como terminarão as investigações na Europa e no Reino Unido e o processo ambiental no Mississippi. E se a integração vai gerar vantagem duradoura ou concentrar riscos demais em uma só empresa.

O que mudou / O que significa / O que fazer agora

O que mudou: entre março de 2025 e agosto de 2026, uma rede social, um laboratório de IA, uma empresa de foguetes e satélites e uma ferramenta de programação passaram a fazer parte de um mesmo grupo, hoje negociado em bolsa e controlado por Musk.

O que significa: a disputa pela IA não é só entre modelos. É também entre cadeias completas, que vão da distribuição aos chips, à energia e à conexão. Quem tem mais camadas pode avançar mais rápido, mas também concentra mais risco, inclusive regulatório e ambiental.

O que fazer agora: ao avaliar o Grok, ou qualquer ferramenta, olhe para a empresa por trás e para as camadas de que ela depende. Separe sempre o que está em documentos públicos do que é promessa. E desconfie tanto de quem já declarou a vitória quanto de quem já decretou o fracasso. A integração é fato; o domínio futuro é hipótese. O livro separará um do outro.

AÇÃO EM DEZ MINUTOS

Escolha uma ferramenta de IA que você usa e desenhe no seu Diário de Fronteira a "pilha" dela, de cima para baixo: onde você a encontra (aplicativo, rede social, carro, celular), qual empresa faz o modelo, onde ficam os data centers, como os dados chegam até você e quem paga a conta. Marque com um ponto de interrogação cada camada que você não sabe. Essas são as perguntas que valem a pena acompanhar.

Musk tenta construir um império que vai do chip ao satélite. Mas, se olharmos para trás, o endereço simbólico do futuro já mudou de lugar várias vezes: da IBM para a Microsoft, da Microsoft para o Google, e agora para os laboratórios de fronteira. No próximo capítulo, vou contar essa mudança de dentro, a partir do que vivi: de IBM e Google aos frontier labs.

Notas

1. Reuters. "SpaceX acquires xAI in record-setting deal as Musk looks to unify AI and space ambitions." 2 fev. 2026. https://www.reuters.com/business/musks-spacex-merge-with-xai-combined-valuation-125-trillion-bloomberg-news-2026-02-02/; CNBC. "Musk's xAI, SpaceX combo is the biggest merger of all time, valued at $1.25 trillion." 3 fev. 2026. https://www.cnbc.com/2026/02/03/musk-xai-spacex-biggest-merger-ever.html

2. U.S. Securities and Exchange Commission. Space Exploration Technologies Corp., Form S-1/A (Emenda n.º 2), 3 jun. 2026 (fusão com o X em 28 mar. 2025; fusão com a xAI efetiva em 2 fev. 2026; poder de voto de Musk de cerca de 84% após a oferta, segundo o documento). https://www.sec.gov/Archives/edgar/data/1181412/000162828026040364/spaceexplorationtechnologib.htm

3. Korosec, Kirsten. "Tesla to invest $2B in Elon Musk's xAI." TechCrunch, 28 jan. 2026. https://techcrunch.com/2026/01/28/tesla-invested-2b-in-elon-musks-xai/

4. Wang, Echo. "Musk's SpaceX prices record $75 billion IPO at $135 a share." Reuters, 11 jun. 2026. https://www.reuters.com/world/musks-spacex-prices-record-75-billion-ipo-135-share-2026-06-11/; SpaceX. "Space Exploration Technologies Corp. Announces Closing of Initial Public Offering, Including Full Exercise of Underwriters' Option to Purchase Additional Shares." Jun. 2026. https://ir.spacex.com/updates/releases-details/2026/Space-Exploration-Technologies-Corp--Announces-Closing-of-Initial-Public-Offering-Including-Full-Exercise-of-Underwriters-Option-to-Purchase-Additional-Shares-2026-RgoR-Y1Vwh/default.aspx

5. Ha, Anthony. "SpaceX officially closes its Cursor acquisition." TechCrunch, 15 ago. 2026. https://techcrunch.com/2026/08/15/spacex-officially-closes-its-cursor-acquisition/; Cursor. "Cursor is now a part of SpaceX." Ago. 2026. https://cursor.com/blog/joining-spacex

6. SEC. Space Exploration Technologies Corp., Form S-1/A (Emenda n.º 2), 3 jun. 2026, seções "Prospectus Summary" e "Business" (usuários ativos mensais em 31 mar. 2026; afirmação sobre dados em tempo real). Mesmo URL da nota 2.

7. SpaceX. "SpaceX Reports Second Quarter 2026 Results." Ago. 2026 (lançamento do Grok 4.5 em julho, treinado com a Cursor). https://ir.spacex.com/files/doc_financials/2026/q2/SpaceX-Reports-Second-Quarter-2026-Results.pdf; SpaceXAI. "Introducing Grok 4.7." Set. 2026. https://x.ai/news/grok-4-7

8. SEC. Form S-1/A (Emenda n.º 2), 3 jun. 2026 (definições de Colossus e Colossus II; contratos com a Anthropic de maio de 2026). Mesmo URL da nota 2; Reuters. "Musk says SpaceX agreed only six-month Colossus AI lease to Anthropic." 28 maio 2026. https://www.reuters.com/technology/musk-says-spacex-did-not-commit-long-term-colossus-lease-with-anthropic-2026-05-28/

9. SpaceX. "SpaceX Reports Second Quarter 2026 Results." Ago. 2026. https://www.sec.gov/Archives/edgar/data/1181412/000162828026052515/earningsreleaseq22608042.htm; The Motley Fool. "SpaceX (SPCX) Q2 2026 Earnings Call Transcript." 11 ago. 2026 (meta de mais de 2 GW no fim do ano). https://www.fool.com/earnings/call-transcripts/2026/08/11/spacex-spcx-q2-2026-earnings-call-transcript/

10. SEC. Form S-1/A (Emenda n.º 2), 3 jun. 2026 (resultados por segmento em 2025; cerca de 9 mil satélites de banda larga e 164 países, territórios e outros mercados em 31 mar. 2026). Mesmo URL da nota 2.

11. SEC. Form S-1/A (Emenda n.º 2), 3 jun. 2026 (computação de IA orbital "as early as 2028"; Terafab com Tesla e Intel). Mesmo URL da nota 2.

12. Saeedy, Alexander; Driebusch, Corrie; Peterson, Becky; Maidenberg, Micah. "Inside Elon Musk's $1.25 Trillion AI and Space Megamerger." The Wall Street Journal, 5 fev. 2026. https://www.wsj.com/tech/elon-musk-xai-spacex-merger-2896ae1e

13. Turyshev, Slava G. "Orbital Data Centers: Spacecraft Constraints and Economic Viability." arXiv:2604.27197, versão de 24 ago. 2026. https://arxiv.org/abs/2604.27197

14. The Guardian. "Elon Musk's AI firm apologizes after chatbot Grok praises Hitler." 12 jul. 2025. https://www.theguardian.com/us-news/2025/jul/12/elon-musk-grok-antisemitic

15. Ofcom. "Ofcom launches investigation into X over Grok sexualised imagery." 12 jan. 2026. https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/ofcom-launches-investigation-into-x-over-grok-sexualised-imagery; Reuters. "Musk dealt blow over Grok deepfakes, but regulatory fight far from over." 15 jan. 2026. https://www.reuters.com/legal/government/musk-dealt-blow-over-grok-deepfakes-regulatory-fight-far-over-2026-01-15/

16. European Commission. "Commission investigates Grok and X's recommender systems under the Digital Services Act." IP/26/203, 26 jan. 2026. https://ec.europa.eu/commission/presscorner/api/files/document/print/en/ip_26_203/IP_26_203_EN.pdf

17. SEC. Form S-1/A (Emenda n.º 2), 3 jun. 2026, seções "Risk Factors" e "Legal Proceedings" (investigações e ações sobre imagens geradas pelo Grok; multa de 120 milhões de euros aplicada pela Comissão Europeia em 5 dez. 2025, contestada em 16 fev. 2026). Mesmo URL da nota 2.

18. Earthjustice. "NAACP Sues xAI for Illegal Pollution from Data Center Power Plant." 14 abr. 2026. https://earthjustice.org/press/2026/xai-sued-for-illegal-power-plant; AI Lawsuit Tracker. "NAACP v. xAI Corp." Revisado em 28 set. 2026. https://ailawsuittracker.com/cases/naacp-v-xai-corp-3-26-cv-00074/

CAPÍTULO 14

De IBM e Google aos frontier labs

Tese do capítulo: o endereço simbólico do futuro muda antes de os antigos gigantes desaparecerem.

No capítulo anterior, vimos Musk tentar juntar todas as camadas da IA dentro de um só grupo. Mas, antes de qualquer empresa controlar uma pilha, ela precisa atrair as pessoas que vão construí-la. E as pessoas mudam de endereço antes do dinheiro, antes das manchetes e muito antes dos balanços.

Eu sei disso porque já trabalhei dentro de um desses endereços. Na IBM Brasil, tornei-me a primeira mulher Diretora Nacional de Vendas da linha PSG, de computadores pessoais, para o varejo, e trabalhei com a linha Aptiva, numa época em que o nome IBM era quase sinônimo de computador. Ter aquele crachá dizia alguma coisa sobre quem você era.

Fui responsável por colocar o primeiro computador pessoal da IBM no varejo, em lojas como Walmart, Ponto Frio e Carrefour. Lançamos a linha Aptiva no Brasil com recordes de vendas e de rentabilidade, e a IBM patrocinou o meu primeiro MBA. Mas não foi um sentimento mágico que me sustentou naquele cargo. Foram evidências: contas que eu tinha aberto, problemas que eu tinha resolvido, salas que eu já tinha atravessado. A IBM me ensinou a ler uma sala.

Hoje, quando converso com jovens talentosos, o crachá dos sonhos é outro. Por muitos anos foi o do Google. Agora, para muita gente, é o de um laboratório de fronteira, como OpenAI, Anthropic ou xAI, ou o da própria startup. Este capítulo segue essa mudança de endereço, e o que ela diz e não diz sobre os gigantes que ficam para trás.

1. Dentro da revolução do PC: Microsoft, varejo, IBM Aptiva e o computador chegando às famílias

No Capítulo 1, contei que trabalhei com o Microsoft Office, ajudei a levar computadores para os supermercados e, depois, trabalhei com a Aptiva na IBM. Em todas essas fases, o trabalho era o mesmo: convencer famílias de que aquela máquina tinha lugar dentro de casa.

A Aptiva nasceu exatamente para isso. A IBM lançou a linha em setembro de 1994, como sucessora do PS/1, para o mercado doméstico.[1] Em outubro daquele ano, a empresa reorganizou seus computadores pessoais em quatro linhas, e a Aptiva era a linha voltada para casa.[2] Os computadores eram vendidos em pacotes, com monitor, alto-falantes, teclado e mouse: tudo o que uma família precisava para começar.[1]

Por trás da marca forte, porém, a IBM vivia seu momento mais difícil. Em 1993, registrou prejuízo líquido de US$ 8,1 bilhões, e Lou Gerstner, vindo de fora da empresa, assumiu o comando naquele ano. Em 1994, a Compaq, a primeira empresa a copiar com sucesso o PC da IBM, tomou dela o posto de maior fabricante de computadores pessoais do mundo.[3] A empresa que tinha popularizado o computador pessoal estava perdendo a corrida que ela mesma tinha começado.

A IBM tinha prestígio, tecnologia e uma marca que todo mundo conhecia. O que ela não controlava mais era a imaginação do mercado sobre para onde o computador estava indo.

NÃO CONFUNDA

Big tech é a grande empresa de tecnologia estabelecida, como Google, Microsoft, Meta, Apple ou Amazon. Pode ter laboratórios de IA de ponta dentro de casa.

Frontier lab, ou laboratório de fronteira, é a organização cuja missão central é construir os modelos mais avançados, como OpenAI, Anthropic ou xAI. O Google DeepMind é um laboratório de fronteira dentro de uma big tech.

Startup de IA é a empresa nova que constrói produtos, ou até modelos próprios, muitas vezes fundada por quem saiu dos dois primeiros.

2. Do sonho de trabalhar no Google ao desejo de entrar em OpenAI, Anthropic, xAI ou fundar uma startup

Por mais de uma década, o Google foi o endereço do futuro. Em 2017, a revista Fortune colocou a empresa em primeiro lugar na sua lista das melhores empresas para trabalhar nos Estados Unidos pela oitava vez em onze anos.[4] Para uma geração de engenheiros, entrar no Google era o equivalente ao que entrar na IBM tinha sido para a minha geração.

A mudança ficou visível em 2025, quando o talento em IA virou objeto de leilão. Em junho daquele ano, Sam Altman disse que a Meta tinha oferecido bônus de contratação de até US$ 100 milhões para atrair funcionários da OpenAI para seu novo laboratório de "superinteligência".[5] Em julho, a Thinking

Machines Lab, fundada por Mira Murati, ex-diretora de tecnologia da OpenAI, levantou US$ 2 bilhões em sua primeira rodada, com valor de US$ 12 bilhões, antes de lançar seu primeiro produto.[6]

Em 2026, o movimento chegou ao coração do Google. No Capítulo 10, vimos que Noam Shazeer anunciou a ida para a OpenAI e John Jumper, para a Anthropic, e que, em 5 de agosto, Jeff Dean, Sanjay Ghemawat, Oriol Vinyals e Quoc Le saíram para fundar uma startup.[7] A empresa se chama Discovery Loop. Foi criada como uma *public benefit corporation*, um tipo de empresa americana que se compromete com uma missão pública além do lucro, e quer automatizar a pesquisa em aprendizado de máquina, ciência e engenharia. O detalhe que mais me chamou a atenção: a própria Alphabet entrou como investidora e fechou com a nova empresa uma parceria de longo prazo em nuvem e computação.[8] Em setembro, segundo o Business Insider, com base em fontes, a Discovery Loop já buscava nova rodada com valor de cerca de US$ 50 bilhões.[9]

O mesmo deslocamento aparece nos dados, e não só nos nomes famosos. Segundo o relatório de talentos de 2026 da SignalFire, gestora de capital de risco que acompanha mais de 650 milhões de perfis profissionais, os formandos de 2025 dos 20 melhores cursos de computação dos Estados Unidos tinham duas vezes mais chance de se declarar fundadores do que a turma de 2022, e 45% menos chance de ocupar uma vaga de engenharia em uma das grandes empresas de tecnologia. Parte disso é escolha. Parte é falta de opção: as contratações de recém-formados nessas empresas caíram cerca de 65% em relação a 2019.[10] Vamos voltar a esse degrau quebrado da carreira mais adiante no livro.

Há uma ironia no mesmo relatório. Os laboratórios de fronteira estão ficando parecidos com as big techs que pretendem substituir. Segundo a SignalFire, OpenAI, Anthropic, Google

DeepMind, xAI, Mistral e Cohere já têm, proporcionalmente, mais gente em recrutamento, finanças e jurídico do que os gigantes tradicionais, e montam estruturas de vendas e marketing parecidas com as deles.[10] O laboratório rebelde de hoje está aprendendo a ser a grande empresa de amanhã.

O talento vota com os pés antes de o mercado votar com o dinheiro.

3. Como aconteceu com a IBM: perder o monopólio do imaginário não é o mesmo que deixar de ser uma grande empresa

A história da IBM tem um final que muita gente esquece. Em maio de 2001, a empresa retirou do mercado a última Aptiva e saiu do varejo doméstico.[1] Em 2005, concluiu a venda de toda a sua divisão de computadores pessoais para a chinesa Lenovo, por US$ 1,25 bilhão, ficando com cerca de 19% da compradora.[11] Para quem viveu o PC por dentro, parecia o fim de uma era. E era. Mas não era o fim da IBM.

Em 2025, a IBM faturou US$ 67,5 bilhões. Segundo a empresa, o software já representava cerca de 45% da receita, o caixa livre foi o maior em mais de uma década, e os negócios ligados à IA generativa somavam mais de US$ 12,5 bilhões desde o início.[12] A IBM deixou de ser o lugar onde os jovens sonhavam trabalhar, mas continuou sendo uma grande empresa, reinventada em torno de software, nuvem híbrida e serviços para corporações.

O Google está muito longe de ser a IBM dos anos 1990. Em 2025, a Alphabet passou pela primeira vez dos US$ 400 bilhões de receita anual.[13] No Capítulo 10, vimos que o Gemini cresce, que o Google tem distribuição gigantesca e um dos melhores

laboratórios do mundo. O que mudou não foi o tamanho. Foi o símbolo. O Google já não tem, sozinho, o lugar de "onde o futuro acontece".

E há uma diferença importante em relação à minha época. A IBM perdeu o PC para concorrentes que não dependiam dela. O Google, ao contrário, investe em quem sai e fornece computação a quem sai, como fez com a Discovery Loop.[8] Os gigantes de hoje aprenderam a lucrar até com a própria perda de prestígio. Isso pode ser sabedoria. Também pode ser a forma moderna de continuar no centro sem ser mais o centro do imaginário.

FOTOGRAFIA DA FRONTEIRA — PARA ONDE O TALENTO FOI (6 DE OUTUBRO DE 2026)

Saídas do Google em 2026: Noam Shazeer (OpenAI) e John Jumper (Anthropic), em junho; Jeff Dean, Sanjay Ghemawat, Oriol Vinyals e Quoc Le fundam a Discovery Loop (5 ago.), com a Alphabet como investidora.

Formandos de elite (SignalFire, jun. 2026): duas vezes mais propensos a se declarar fundadores do que em 2022; 45% menos propensos a uma vaga de engenharia numa big tech.

Gigantes: Alphabet com mais de US$ 400 bilhões de receita em 2025; IBM com US$ 67,5 bilhões.

Fatos vivos: confira o Atlas Vivo da IA.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: é como o clube onde todo mundo sonhava jogar. Ele continua grande, rico e cheio de títulos, mas os melhores jogadores jovens agora querem ir para outro time, ou montar o próprio. O clube antigo não acabou. Só deixou de ser o único sonho.

Para quem tem 65 anos: lembre de quando trabalhar no Banco do Brasil, na Petrobras ou numa multinacional famosa era o auge de uma carreira. Essas instituições continuam importantes, mas muitos jovens hoje sonham em abrir o próprio negócio. A régua do prestígio mudou de lugar.

O QUE AINDA NÃO SABEMOS

Se os laboratórios de fronteira vão manter o magnetismo quando ficarem tão grandes e burocráticos quanto as big techs. Se startups como a Discovery Loop vão justificar valores tão altos antes de ter produto. Se o fechamento das vagas de entrada vai empurrar mais jovens para empreender ou deixar uma geração sem degrau. E se o Google vai recuperar o símbolo do futuro ou se acostumar a dividi-lo.

O que mudou / O que significa / O que fazer agora

O que mudou: o endereço dos sonhos da tecnologia saiu de um gigante para vários lugares ao mesmo tempo: os laboratórios de fronteira e as startups fundadas por quem sai deles. Em 2025 e 2026, salários recordes, rodadas bilionárias e saídas de nomes históricos do Google tornaram essa mudança visível.

O que significa: perder o monopólio do imaginário não é o mesmo que perder relevância. A IBM continua grande, e o Google,

muito maior. Mas quem atrai as pessoas mais ambiciosas costuma desenhar a próxima geração de produtos, e os gigantes passam a disputar esse futuro como investidores, fornecedores e parceiros, não apenas como protagonistas.

O que fazer agora: quando você ler sobre uma empresa de IA, não olhe só para a receita, o valor de mercado ou o último modelo. Olhe também para quem está chegando e quem está saindo. Se você é jovem, pense no que cada endereço ensina, não só no crachá. Se você lidera uma empresa, pergunte por que alguém talentoso escolheria trabalhar com você hoje. Preste atenção não apenas ao dinheiro, mas ao lugar para onde as pessoas mais ambiciosas acreditam que o futuro foi.

AÇÃO EM DEZ MINUTOS

No seu Diário de Fronteira, escreva o nome da empresa onde você sonhava trabalhar aos 20 anos. Ao lado, escreva onde um jovem talentoso da sua família ou da sua equipe sonha trabalhar hoje. Se não souber, pergunte a ele nesta semana. Depois, anote uma frase: o que mudou entre os dois sonhos, e o que isso diz sobre o lugar onde o futuro está sendo construído?

Mas talento é só metade da equação. Para pagar esses salários, esses chips e esses data centers, os laboratórios de fronteira precisam de quantias de dinheiro que nenhuma startup jamais levantou. No próximo capítulo, vamos entender a linguagem desse dinheiro: receita, ARR, valuation, rodadas, S-1 e IPO.

Notas

1. HomeComputerMuseum. "IBM Aptiva." Consultado em 6 out. 2026 (lançamento em setembro de 1994 como sucessora do PS/1; vendida em pacotes; última unidade retirada em maio de 2001, quando a IBM saiu do

mercado doméstico).
https://www.homecomputermuseum.nl/en/collectie/ibm/ibm-aptiva/

2. The New York Times. "I.B.M. Plans to Cut PC Lines to 4 From 9." 17 out. 1994. https://www.nytimes.com/1994/10/17/business/ibm-plans-to-cut-pc-lines-to-4-from-9.html

3. CNET. "IBM's outsider: A look back at Lou." Publicado após os resultados de 2001; consultado em 6 out. 2026. https://www.cnet.com/tech/tech-industry/ibms-outsider-a-look-back-at-lou/

4. Fortune. "Fortune 100 Best Companies to Work For 2017." Mar. 2017. https://fortune.com/ranking/best-companies/2017/

5. Butts, Dylan. "Sam Altman says Meta offered OpenAI staff $100 million bonuses, as Mark Zuckerberg ramps up AI poaching efforts." CNBC, 18 jun. 2025. https://www.cnbc.com/2025/06/18/sam-altman-says-meta-tried-to-poach-openai-staff-with-100-million-bonuses-mark-zuckerberg.html

6. TechCrunch. "Mira Murati's Thinking Machines Lab is worth $12B in seed round." 15 jul. 2025. https://techcrunch.com/2025/07/15/mira-muratis-thinking-machines-lab-is-worth-12b-in-seed-round/

7. Reuters. "US scientist John Jumper to leave Google DeepMind for Anthropic." 19 jun. 2026. https://www.reuters.com/technology/us-scientist-john-jumper-leave-google-deepmind-anthropic-2026-06-19/; Whitwam, Ryan. "Google's AI shake-up: DeepMind's Hassabis steps aside, senior scientists depart." Ars Technica, 5 ago. 2026. https://arstechnica.com/gadgets/2026/08/googles-ai-shakeup-deepminds-hassabis-steps-aside-senior-scientists-depart/

8. Wilson Sonsini Goodrich & Rosati. "Wilson Sonsini Advises Discovery Loop on Launch and Initial Funding." 5 ago. 2026. https://www.wsgr.com/en/insights/wilson-sonsini-advises-discovery-loop-on-launch-and-initial-funding.html; Jacobs, Jordan. "Radical Reads: Jeff Dean on Launching Discovery Loop." Radical Ventures, 9 ago. 2026. https://radical.vc/radical-reads-jeff-dean-on-launching-discovery-loop/

9. Business Insider. "Former Google chief scientist Jeff Dean is raising again for his new AI startup, seeking around a $50 billion valuation." Set. 2026.

https://www.businessinsider.com/jeff-deans-startup-discovery-loop-is-eyeing-a-valuation-2026-9

10. SignalFire. "SignalFire's State of Talent Report - 2026." 22 jun. 2026. https://www.signalfire.com/blog/signalfire-state-of-talent-report-2026

11. Lenovo. "Lenovo Completes Acquisition of IBM's Personal Computing Division." 1 maio 2005. https://news.lenovo.com/pressroom/press-releases/lenovo-completes-acquisition-ibms-personal-computing-division/

12. IBM. "IBM Releases Fourth-Quarter Results." 28 jan. 2026. https://newsroom.ibm.com/2026-01-28-IBM-RELEASES-FOURTH-QUARTER-RESULTS; IBM. Relatório anual 2025 ("Let's Create Smarter Business"). https://www.ibm.com/downloads/documents/us-en/15db52348fc203a4

13. Alphabet. "Alphabet Announces Fourth Quarter and Fiscal Year 2025 Results." Fev. 2026. https://s206.q4cdn.com/479360582/files/doc_financials/2025/q4/2025q4-alphabet-earnings-release.pdf

CAPÍTULO 15

IPOs, valuations e a corrida por capital

Tese do capítulo: a fronteira exige tanto capital que sua estrutura financeira se torna parte da história tecnológica.

Terminei o capítulo anterior dizendo que talento é só metade da equação. A outra metade é dinheiro, e não qualquer dinheiro. Para pagar pesquisadores, chips e data centers, os laboratórios de fronteira passaram a levantar quantias que, poucos anos atrás, pareceriam ficção.

E, junto com o dinheiro, chegou uma linguagem nova para muita gente. Receita, ARR, valuation, rodada, S-1, IPO. Já vi pessoas muito inteligentes lerem "empresa avaliada em US$ 900 bilhões" e entenderem que a empresa tinha US$ 900 bilhões no banco. Ou lerem "ARR de US$ 40 bilhões" e acharem que era lucro. Não é culpa delas. As manchetes misturam tudo.

Este capítulo é um pequeno dicionário com exemplos reais. Porque, em 2026, a estrutura financeira da IA virou parte da própria história da tecnologia.

1. Receita, ARR, valuation, rodada, S-1 e IPO: o que cada palavra significa

Receita é o que a empresa faturou em um período: um trimestre, um ano. É o número mais concreto. Segundo o prospecto da

Anthropic visto pela Reuters, a receita da empresa em 2025 foi de quase US$ 4,6 bilhões, doze vezes mais que no ano anterior.[1] A OpenAI, segundo o Wall Street Journal, informou a investidores uma receita de US$ 6,7 bilhões no segundo trimestre de 2026.[2]

ARR, ou receita recorrente anualizada, é uma fotografia multiplicada. Se a empresa recebe US$ 10 milhões por mês de forma recorrente, pode dizer que tem US$ 120 milhões de ARR. É útil para medir velocidade, mas não é receita de um ano inteiro, nem lucro, nem caixa. Veja a diferença na prática. Em maio de 2026, a Anthropic informou que seu ritmo anualizado de receita tinha passado de US$ 47 bilhões. No fim de julho, segundo a CNBC, disse a investidores que o ritmo chegara a US$ 65 bilhões.[2,3] Mas a receita que ela de fato registrou em 2025 inteiro foi de US$ 4,6 bilhões.[1] As duas coisas podem ser verdadeiras ao mesmo tempo: uma mede o passado, a outra mede a velocidade do momento, projetada para a frente. No caso da OpenAI, a CNBC confirmou um ritmo anualizado acima de US$ 40 bilhões em agosto.[4]

Lucro e caixa são outra conversa. Uma empresa pode faturar bilhões e perder ainda mais. Segundo o mesmo prospecto, a Anthropic teve prejuízo líquido de cerca de US$ 42 bilhões em 2025. Desse total, cerca de US$ 34 bilhões foram um ajuste contábil, ligado ao valor estimado de financiamentos que podem virar ações, e não dinheiro gasto na operação. Ainda assim, o prejuízo operacional passou de US$ 8 bilhões. E o caixa, somado a aplicações de curto prazo, era de US$ 20,28 bilhões no fim do ano.[1] Faturamento, prejuízo e caixa são três números diferentes.

Rodada e valuation. Uma rodada é quando investidores colocam dinheiro na empresa em troca de participação. O valuation é o preço implícito da empresa inteira naquela negociação. Em 31 de março de 2026, a OpenAI fechou uma

rodada de US$ 122 bilhões em capital comprometido, com valuation de US$ 852 bilhões depois do aporte.[5] "Comprometido" importa: segundo a Bloomberg, US$ 35 bilhões dos US$ 50 bilhões da Amazon dependiam de a OpenAI abrir o capital ou atingir o marco tecnológico da inteligência artificial geral.[6] Em 28 de maio, a Anthropic anunciou US$ 65 bilhões a um valuation de US$ 965 bilhões, superando pela primeira vez a OpenAI, incluindo US$ 15 bilhões já prometidos antes por grandes provedores de nuvem.[7] Um valuation de quase US$ 1 trilhão não significa US$ 1 trilhão no banco. Significa que alguém aceitou pagar por uma fatia um preço que, multiplicado pela empresa inteira, dá esse número.

S-1 e IPO. O S-1 é o documento de registro que uma empresa entrega à SEC, a comissão de valores mobiliários dos Estados Unidos, para vender ações ao público. Desde 2017, qualquer empresa pode enviar primeiro um rascunho confidencial para análise. Mas precisa torná-lo público pelo menos 15 dias antes de começar a apresentar a oferta aos investidores, a chamada *roadshow*.[8] A IPO, a oferta pública inicial, é o momento em que as ações passam a ser vendidas e negociadas em bolsa.

NÃO CONFUNDA

ARR não é receita do ano, lucro nem caixa: é o ritmo do momento, anualizado.

Valuation não é dinheiro no banco: é o preço implícito da empresa inteira numa negociação.

Capital comprometido não é dinheiro já recebido: pode ter condições.

S-1 confidencial não é IPO marcada: é a opção de abrir o capital, se e quando a empresa quiser.

2. OpenAI e Anthropic apresentaram S-1 confidencial: isso cria opção, não data garantida

Em 1.º de junho de 2026, a Anthropic anunciou que tinha enviado à SEC um rascunho confidencial de S-1. A própria empresa escreveu que isso lhe dava "a opção de abrir o capital" depois da análise da SEC, e que a oferta dependeria das condições de mercado.[9] Em 8 de junho, a OpenAI anunciou o mesmo passo, com uma franqueza rara: "Ainda não decidimos o momento; pode demorar".[10]

A SpaceX, que vimos no Capítulo 13, mostrou o caminho completo. Tornou público o seu S-1 em maio, fixou o preço em 11 de junho, a US$ 135 por ação, e estreou no dia seguinte, avaliada em cerca de US$ 1,77 trilhão. No primeiro dia, as ações subiram 19%, para US$ 160. No fim de setembro, valiam cerca de US$ 147.[1] É uma boa lição: depois da IPO, o valuation muda todos os dias, com o humor do mercado.

Com a OpenAI e a Anthropic, a opção ainda não virou data. Em 19 de agosto, a diretora financeira da OpenAI, Sarah Friar, disse aos funcionários que a empresa "será uma empresa aberta em 2027", ou antes, se o negócio continuasse acelerando.[2] Em 12 de setembro, Sam Altman disse que a OpenAI não abriria o capital em 2026, por causa das preocupações com segurança.[11] E, em 29 de setembro, na conferência da empresa para desenvolvedores, afirmou não ter "um prazo específico em mente".[12]

A Anthropic, segundo a Reuters, chegou a mirar uma estreia em meados de outubro.[13] Depois, a mesma agência informou, com base em fontes, que a oferta provavelmente ficaria para depois das eleições americanas de novembro, e que a empresa poderia buscar

um valor acima de US$ 2 trilhões.[1] No fim de setembro, a Reuters teve acesso ao prospecto, que ainda não tinha sido tornado público.[1] Até 6 de outubro de 2026, nenhuma das duas tinha anunciado data de oferta.

FOTOGRAFIA DA FRONTEIRA — A CORRIDA POR CAPITAL (6 DE OUTUBRO DE 2026)

OpenAI: rodada de US$ 122 bilhões a US$ 852 bilhões (31 mar.); S-1 confidencial (anunciado em 8 jun.); IPO descartada para 2026, sem data.

Anthropic: rodada de US$ 65 bilhões a US$ 965 bilhões (28 maio); S-1 confidencial (1.º jun.); estreia esperada, segundo a Reuters, para depois das eleições de novembro, sem data anunciada.

SpaceX: na bolsa desde 12 jun. (SPCX), estreia a US$ 135 por ação.

Fatos vivos: confira o Atlas Vivo da IA.

3. Por que empresas extraordinárias podem crescer rapidamente, gastar ainda mais e buscar o mercado público

Se a Anthropic multiplicou a receita por doze, por que perdeu tanto dinheiro? Porque a fronteira é cara. Em 2025, a empresa gastou US$ 7,33 bilhões com computação e infraestrutura, mais da metade dos seus US$ 12,65 bilhões de despesas operacionais. E, segundo o prospecto, tem US$ 518 bilhões em compromissos com nuvem, computação e infraestrutura para os próximos anos.[1] No Capítulo 13, vimos que ela aceitou pagar à SpaceX US$ 1,25 bilhão por mês por acesso a GPUs. No Capítulo 12, vimos as grandes plataformas americanas projetarem mais de US$ 735 bilhões de investimentos em 2026.

É uma lógica parecida com a do varejo, só que em outra escala: quem cresce rápido precisa de capital de giro antes de receber. Na IA, o "estoque" é a computação, comprada com anos de antecedência, para um mercado que ainda está se formando.

Na American Photonics, vivi o lado bom e o lado pesado do crescimento. A empresa crescia tanto e tão rápido que começou a ficar maior do que nós. Steve e eu éramos o conselho inteiro, só nós dois carregando todas as decisões, e a empresa pedia o que duas pessoas não conseguem oferecer sozinhas: processos de verdade, investimento sério em máquinas, mais mãos, mais mão de obra. Em 2025, tomei uma das decisões mais difíceis da minha vida profissional: colocar a empresa à venda. Não porque estivesse falhando. Pelo contrário: ela merecia mais pista do que nós dois conseguíamos construir.

Por isso o mercado público se torna atraente. Friar resumiu assim para os funcionários: a IPO "não é uma linha de chegada", é "um marco, mais uma captação".[2] A bolsa é a maior fonte de capital que existe. Também permite que funcionários e investidores antigos vendam parte das suas ações.

Mas a bolsa cobra um preço: transparência e pressão de curto prazo. No prospecto, a Anthropic informou que quase um quarto da sua receita de 2025 veio de dois clientes e que muitos dos maiores clientes não têm contratos de longo prazo e podem reduzir os gastos.[1] Altman, por sua vez, disse temer "decepcionar os apoiadores de Wall Street" em nome da segurança.[12] Quem abre o capital passa a prestar contas a cada trimestre, inclusive sobre decisões de não lançar um produto.

Valor de mercado é opinião com preço. Caixa é fato com prazo.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: imagine que alguém compra 1% da sua barraca de limonada por R$ 100. Na manchete, sua barraca "vale R$ 10 mil". Mas no seu bolso há R$ 100, e você ainda precisa comprar os limões. ARR é dizer: "Hoje vendi R$ 30; em um ano, seriam R$ 10.950". Talvez. Se chover a semana inteira, não.

Para quem tem 65 anos: é como a diferença entre o valor de avaliação de um imóvel e o dinheiro que entra com o aluguel. O imóvel pode valer muito no papel e ainda assim você precisar pagar o condomínio todo mês. E, para vender, é preciso alguém disposto a pagar aquele preço no dia.

O QUE AINDA NÃO SABEMOS

Quando OpenAI e Anthropic vão tornar públicos seus prospectos e se as estreias ficarão para 2026 ou 2027. Se os ritmos anualizados de receita vão se transformar em receita anual na mesma proporção. Quando essas empresas terão lucro operacional. Se o mercado vai pagar os valuations esperados, ou se a queda recente das ações de IA vai baixar os preços. E quanto dos compromissos bilionários com computação será, de fato, usado.

O que mudou / O que significa / O que fazer agora

O que mudou: em 2026, os laboratórios de fronteira passaram a ser avaliados perto de US$ 1 trilhão em rodadas privadas, e a corrida chegou à bolsa: a SpaceX estreou em junho, e a OpenAI e a Anthropic entregaram rascunhos confidenciais de S-1, sem data definida para a oferta.

O que significa: a IA de fronteira virou uma indústria de capital intensivo, como energia ou telecomunicações. A receita cresce rápido, mas os custos com computação crescem junto, e a estrutura financeira passou a definir quem consegue continuar na corrida.

O que fazer agora: quando ler uma manchete financeira sobre IA, faça três perguntas. Esse número é receita, ritmo anualizado, valuation ou caixa? Quem o divulgou: a empresa, um documento oficial ou uma fonte anônima? E há data confirmada ou apenas uma opção? Uma IPO não prova que a inteligência é melhor; mostra que a corrida entrou na escala do mercado de capitais.

AÇÃO EM DEZ MINUTOS

Procure uma notícia financeira recente sobre uma empresa de IA. No seu Diário de Fronteira, copie o número principal e classifique-o: receita, ARR, lucro ou prejuízo, caixa, valuation ou capital comprometido. Anote quem divulgou o número e a data. Depois, reescreva a manchete em uma frase que um adolescente entenderia sem se enganar.

Dinheiro, porém, não garante o futuro. Houve empresas com caixa, marca e participação de mercado gigantescos que perderam a liderança em poucos anos quando a interface mudou. Eu vi isso de perto. No próximo capítulo: Nokia, BlackBerry e o momento Apple da IA.

Notas

1. Wang, Echo; Saini, Manya. "Exclusive: Anthropic's IPO prospectus shows sweeping AI vision, surging costs." Reuters, 28 set. 2026 (prospecto visto pela Reuters, ainda não público; receita, prejuízo, caixa e compromissos de 2025; adiamento para depois das eleições de novembro; desempenho das

ações da SpaceX). Reprodução distribuída pela Reuters: https://d2233.cms.socastsrm.com/2026/09/28/exclusive-anthropics-ipo-prospectus-shows-sweeping-ai-vision-surging-costs/

2. Capoot, Ashley; Rooney, Kate. "OpenAI 'will be a public company in 2027' or sooner, CFO Friar tells employees." CNBC, 19 ago. 2026 (inclui a receita do 2.º trimestre informada a investidores, segundo o Wall Street Journal, e o ritmo anualizado da Anthropic no fim de julho). https://www.cnbc.com/2026/08/19/open-ai-ipo-timing-2027-friar.html

3. Anthropic. "Anthropic raises $65B in Series H funding at $965B post-money valuation." 28 maio 2026. https://www.anthropic.com/news/series-h

4. Rooney, Kate. "OpenAI CFO Friar tells investors that enterprise business now bigger than consumer by revenue." CNBC, 14 ago. 2026. https://www.cnbc.com/2026/08/14/openai-cfo-friar-tells-investors-that-enterprise-bigger-than-consumer.html

5. OpenAI. "OpenAI raises $122 billion to accelerate the next phase of AI." 31 mar. 2026. https://openai.com/index/accelerating-the-next-phase-ai/

6. Ghaffary, Shirin. "OpenAI Valued at $852 Billion After Completing $122 Billion Round." Bloomberg, 31 mar. 2026. https://www.bloomberg.com/news/articles/2026-03-31/openai-valued-at-852-billion-after-completing-122-billion-round

7. Anthropic, nota 3; Reuters. "Anthropic's valuation surges to $965 billion, surpassing OpenAI." 28 maio 2026. https://www.reuters.com/business/anthropic-raises-65-billion-now-valued-965-billion-2026-05-28/

8. U.S. Securities and Exchange Commission. "SEC's Division of Corporation Finance Expands Popular JOBS Act Benefit to All Companies." Comunicado 2017-121, 29 jun. 2017. https://www.sec.gov/newsroom/press-releases/2017-121

9. Anthropic. "Anthropic confidentially submits draft S-1 to the SEC." 1 jun. 2026. https://www.anthropic.com/news/confidential-draft-s1-sec

10. OpenAI. "Confidential submission of draft S-1 to the SEC." 8 jun. 2026. https://openai.com/index/openai-submits-confidential-s-1/

11. Ha, Anthony. "OpenAI's Sam Altman says it would be 'ill-advised' to go public in 2026." TechCrunch, 12 set. 2026. https://techcrunch.com/2026/09/12/openais-sam-altman-says-it-would-be-ill-advised-to-go-public-in-2026/; Baertlein, Lisa. "OpenAI's Altman won't do IPO this year, calls AI extinction risk 'unacceptable'." Reuters, 12 set. 2026. https://www.reuters.com/legal/litigation/openai-ipo-will-not-happen-2026-amid-ai-safety-fears-altman-says-2026-09-12/

12. CNBC. "OpenAI DevDay 2026: Live updates and announcements." 29 set. 2026. https://www.cnbc.com/2026/09/29/openai-devday-2026-live-updates.html; The Verge. "Sam Altman says OpenAI won't go public until its models are safe." 29 set. 2026. https://www.theverge.com/ai-artificial-intelligence/1002505/sam-altman-openai-ipo-devday-ai-safety

13. Reuters. "Exclusive: Anthropic IPO launch shifts toward mid-October, sources say." 4 set. 2026. https://www.reuters.com/world/anthropic-ipo-launch-shifts-toward-mid-october-sources-say-2026-09-04/

CAPÍTULO 16

Nokia, BlackBerry e o momento Apple da IA

Tese do capítulo: líderes podem ser excelentes no jogo antigo e ainda perder quando a interface redefine o produto.

Terminei o capítulo anterior dizendo que dinheiro não garante o futuro. Houve empresas com caixa, marca e participação de mercado gigantescos que perderam a liderança em poucos anos quando a interface mudou. Eu vi isso de perto.

No Capítulo 1, contei que acompanhei Nokia e Motorola em mais de 1.500 pontos de venda e que também fui beta tester do BlackBerry. Nokia e Motorola foram as duas marcas que se revezaram no topo do mundo: a Motorola liderou até 1998, quando a Nokia assumiu o primeiro lugar em vendas de celulares e não o largou por catorze anos.[1] Naquele balcão, a conversa era sobre aparelho, plano e preço. Ninguém perguntava qual era o sistema operacional. Ninguém perguntava quantos aplicativos o telefone tinha.

Antes de entrar na Visual Presence, eu tinha sido beta tester do BlackBerry e da Motorola. Na Visual Presence, uma empresa de serviços de marketing de varejo e de ponto de venda, eu controlava todos os demonstradores da Nokia em nível nacional, mais de 200 pessoas, e todos os lançamentos da Nokia no Brasil. Atendíamos também a Motorola, em mais de 1.500 lojas pelo país. Eu estava

sempre lidando com a tecnologia de ponta que ainda ia ser lançada e que não estava disponível no mercado.

Hoje, quando uso um assistente de IA para escrever, pesquisar e organizar o meu dia, tenho a mesma sensação de estar diante de um balcão que vai mudar de forma. Este capítulo é sobre essa sensação.

1. Nokia e BlackBerry venceram as regras anteriores

É fácil, olhando para trás, tratar Nokia e BlackBerry como empresas que "não entenderam". Não é verdade. Eu vi isso com Nokia e BlackBerry: elas entenderam muito bem o jogo que estava sendo jogado, e venceram.

Em 2009, dois anos depois do lançamento do iPhone, a Nokia ainda vendia 36,4% de todos os celulares do mundo, segundo a Gartner. Entre os smartphones, o Symbian, sistema da Nokia, tinha 46,9% das vendas. O BlackBerry, da canadense Research In Motion, vinha em segundo, com 19,9%. O iPhone era o terceiro, com 14,4%.[2] A Nokia sabia fabricar em escala, distribuir em qualquer país, fazer bateria que durava dias e aparelho que sobrevivia a uma queda no chão. O BlackBerry tinha conquistado executivos e governos com o e-mail no bolso, o teclado físico e a fama de segurança.

As regras daquele jogo eram claras: o produto era o aparelho. Ganhava quem tinha o melhor hardware, o melhor preço para cada faixa de mercado, a melhor relação com as operadoras e a maior presença nas lojas. Por essas regras, a Nokia era quase imbatível.

Até que surgiu um concorrente jogando outro jogo. Em abril de 2007, pouco mais de três meses depois do anúncio do iPhone, o presidente da Microsoft, Steve Ballmer, disse ao USA Today que não havia "nenhuma chance" de o iPhone conquistar uma participação de mercado significativa.[3] Pelas regras antigas, a frase fazia sentido: o aparelho era caro, só funcionava com uma operadora nos Estados Unidos e nem tinha teclado físico.

Em fevereiro de 2011, o então presidente da Nokia, Stephen Elop, mandou a todos os funcionários um memorando que ficou famoso. Disse que a empresa estava sobre uma "plataforma em chamas" e escreveu a frase que resume este capítulo: "A batalha dos aparelhos se tornou uma guerra de ecossistemas." Os concorrentes, dizia ele, não estavam tomando mercado da Nokia com aparelhos, mas com um ecossistema inteiro.[4] Dias depois, a Nokia anunciou que adotaria o Windows Phone, da Microsoft, como sua principal plataforma de smartphones.[5]

O resto aconteceu rápido. No primeiro trimestre de 2012, a Samsung, impulsionada pelos smartphones Galaxy, com o Android do Google, passou a Nokia em vendas de celulares.[1] Em 2014, a Microsoft concluiu a compra da divisão de aparelhos da Nokia, num negócio anunciado por 5,44 bilhões de euros.[6] Em 2016, a BlackBerry anunciou que deixaria de desenvolver os próprios telefones e passaria a terceirizar essa função a parceiros.[7] A Motorola, a antiga líder, foi comprada pelo Google em 2012, num negócio avaliado em US$ 12,5 bilhões, e revendida à Lenovo em 2014 por cerca de US$ 2,91 bilhões.[8]

Mas repare num detalhe que costuma ser esquecido. Nem a Nokia nem a BlackBerry desapareceram. Em 2025, a Nokia faturou 19,9 bilhões de euros, principalmente com equipamentos de rede.[9] Em outubro de 2025, a Nvidia anunciou um investimento de US$ 1 bilhão na empresa, para desenvolver com

ela redes preparadas para a IA.[10] E a BlackBerry, no ano fiscal encerrado em fevereiro de 2026, faturou US$ 549,1 milhões com software de segurança e com o QNX, um sistema que, segundo a empresa, está em mais de 275 milhões de veículos.[11] Elas perderam o produto que o mundo carregava no bolso, não a capacidade de existir.

NÃO CONFUNDA

Aparelho é o objeto que você segura: o telefone, o computador, o fone.

Interface é a forma como você conversa com a tecnologia: o teclado, a tela de toque, a voz, o assistente.

Ecossistema é tudo o que vem junto: sistema operacional, loja, desenvolvedores, pagamentos, serviços e o hábito de quem usa.

2. A Apple transformou o telefone em sistema operacional, loja, câmera, música, internet e identidade

Em 9 de janeiro de 2007, a Apple apresentou o iPhone como três produtos em um: um celular, um iPod com tela larga e controles de toque e um aparelho de internet com e-mail, navegador, busca e mapas. Steve Jobs prometia "a interface mais revolucionária desde o mouse": os dedos.[12]

O passo decisivo veio um ano e meio depois. Em 10 de julho de 2008, a Apple abriu a App Store com mais de 500 aplicativos, mais de 125 deles gratuitos, que podiam ser baixados direto no telefone.[13] Serviços de conteúdo para celular já existiam; no Japão, por exemplo, o i-mode da NTT Docomo oferecia serviços de internet móvel desde 1999.[14] A novidade da Apple foi juntar, num

mesmo lugar, o sistema operacional, a loja, os desenvolvedores, o pagamento e a atualização.

A partir dali, o telefone deixou de ser um aparelho e virou uma plataforma. A câmera virou álbum de família. O iPod virou a trilha sonora do dia. O navegador virou banco, mapa e loja. E o logotipo na parte de trás virou parte da identidade de quem o carregava. O produto não era mais o objeto. Era tudo o que o objeto permitia fazer.

É isso que estou chamando de "momento Apple": o instante em que uma nova interface redefine o que é o produto. Quando isso acontece, as qualidades que davam a vitória no jogo antigo, como bateria, resistência e escala de fábrica, continuam importantes, mas deixam de decidir a partida.

O melhor fabricante do jogo antigo pode perder para quem muda o que significa jogar.

3. Na IA, o vencedor precisará combinar modelo, produto, agentes, dados, distribuição, confiança e preço

O momento Apple da IA ainda não aconteceu por completo. Mas os sinais estão em toda parte, e nenhuma empresa tem, sozinha, todas as peças.

A própria Apple é o melhor exemplo. Ela tem distribuição: mais de 2,5 bilhões de aparelhos ativos, segundo a empresa, em janeiro de 2026.[15] Tem a confiança de muita gente em privacidade. Mas, quando lançou o novo Siri AI, em 14 de setembro de 2026, informou que os novos modelos foram construídos em colaboração com o Google e seus modelos Gemini. O Siri AI

começou em versão de teste, apenas em inglês, com português previsto para outubro, e ficou de fora, inicialmente, do iPhone na União Europeia e da China.[16] A empresa que definiu a última grande interface hoje depende, em parte, do modelo de um concorrente.

A OpenAI tem o caminho inverso. Tem modelo, produto e agentes: no Capítulo 9, vimos os Dots, agentes que ficam sempre ligados. No DevDay de 29 de setembro de 2026, apresentou também extensões que permitem a desenvolvedores construir aplicativos inteiros que "parecem nativos" dentro do ChatGPT, distribuídos pela própria OpenAI.[17] É uma loja de aplicativos nascendo dentro de um assistente. O que a OpenAI ainda não tem é o objeto. Documentos apresentados à Justiça em fevereiro de 2026 indicavam que seu primeiro aparelho, desenhado por Jony Ive, ex-chefe de design da Apple, não chegaria aos clientes antes do fim de fevereiro de 2027.[18] No DevDay, Sam Altman disse não se preocupar em ser o primeiro: "Novos formatos de computação aparecem muito raramente."[17]

O Google, como vimos no Capítulo 10, tem modelo, dados e uma distribuição gigantesca no Android e na Busca. A Microsoft tem o escritório: numa pesquisa da a16z com cem grandes empresas, publicada em janeiro de 2026, o Microsoft 365 Copilot liderava o uso corporativo de assistentes, e 65% das empresas disseram preferir soluções dos fornecedores que já usam, quando disponíveis.[19] A Anthropic, no Capítulo 9, apostou no trabalho profundo e confiável. E o preço virou arma: no mesmo DevDay, a OpenAI destacou a redução de preço do GPT-6.1 Sol.[17]

Há, por isso, bons argumentos para acreditar que não haverá uma única Apple da IA. Na mesma pesquisa, 81% das grandes empresas usavam três ou mais famílias de modelos, contra 68% menos de um ano antes. A liderança variava conforme a tarefa: a

OpenAI à frente na conversa geral e no atendimento, a Anthropic na programação e na análise de dados.[19] A pesquisa tem limites: ouviu cem executivos escolhidos, não uma amostra aleatória, e a própria a16z é investidora da OpenAI. Mas o retrato é de um mercado com vários vencedores, pelo menos por enquanto.

Eu já sinto essa mudança no meu dia. Hoje, em vez de abrir um aplicativo para cada coisa, eu converso com os meus agentes: peço que publiquem um site, organizem os e-mails ou revisem um capítulo deste livro.

FOTOGRAFIA DA FRONTEIRA — QUEM TEM QUAIS PEÇAS (6 DE OUTUBRO DE 2026)

Apple: mais de 2,5 bilhões de aparelhos ativos (jan. 2026); Siri AI lançado em 14 set. 2026, em beta e em inglês, com modelos feitos em colaboração com o Google Gemini.

OpenAI: modelos GPT-6, agentes Dots e extensões de aplicativos dentro do ChatGPT (DevDay, 29 set.); aparelho próprio não esperado antes do fim de fev. 2027.

Empresas (a16z, jan. 2026): 81% usam três ou mais famílias de modelos; 65% preferem o fornecedor que já usam.

Os antigos líderes: Nokia, 19,9 bilhões de euros em redes (2025); BlackBerry, US$ 549,1 milhões em software (ano fiscal 2026).

Fatos vivos: confira o Atlas Vivo da IA.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: imagine que o seu celular, um dia, deixe de ser uma tela cheia de ícones e vire uma conversa. Você pede, e o assistente faz: compra o ingresso, manda a mensagem, organiza o trabalho da escola. Os aplicativos continuam existindo, mas ficam por trás. Quem controlar essa conversa será, para a sua geração, o que o iPhone foi para a geração dos seus pais.

Para quem tem 65 anos: lembre do celular "tijolão" e, depois, do aparelho que cabia no bolso e durava dias. Cada um parecia o fim da linha, até o telefone de tela de toque mudar tudo. Agora a mudança pode ser outra: em vez de aprender onde fica cada botão, você vai poder simplesmente pedir o que precisa, com a sua voz.

O QUE AINDA NÃO SABEMOS

Se a nova interface será um aparelho novo, um assistente dentro do celular que já temos ou as duas coisas. Se a Apple vai recuperar a dianteira com modelos próprios ou continuar dependendo de parceiros. Se a OpenAI conseguirá transformar o ChatGPT numa plataforma tão fechada e lucrativa quanto a App Store. E se as empresas vão continuar usando vários modelos ou, com o tempo, concentrar tudo em um ou dois fornecedores.

O que mudou / O que significa / O que fazer agora

O que mudou: a conversa com a tecnologia está saindo dos ícones e dos aplicativos e indo para assistentes e agentes que entendem pedidos e agem por nós. Em 2026, Apple, OpenAI, Google e Microsoft estão montando, cada uma a partir do que já tem, as peças dessa nova interface.

O que significa: as qualidades que garantem a liderança hoje podem não bastar amanhã. Nokia e BlackBerry não perderam por incompetência; perderam porque o produto mudou de definição. Na IA, ninguém reúne ainda todas as peças, e o mais provável, por enquanto, é que haja vários vencedores, cada um forte em uma camada.

O que fazer agora: olhe para o seu próprio negócio, ou a sua carreira, e pergunte qual parte do que você vende ainda depende da interface antiga. Se o seu cliente pudesse resolver o problema conversando com um assistente, ele ainda passaria pela sua loja, pelo seu site ou pelo seu atendimento? Comece a testar hoje como o seu produto aparece dentro desses assistentes. Talvez não exista uma única Apple da IA — mas certamente existirão empresas que descobrirão tarde demais que o produto mudou.

AÇÃO EM DEZ MINUTOS

No seu Diário de Fronteira, escreva três tarefas que você fez hoje abrindo aplicativos ou sites: pagar uma conta, marcar uma reunião, procurar um produto. Para cada uma, anote se um assistente de IA já poderia fazer isso por você, se faria só em parte ou se ainda não faria. Depois, escolha uma e teste. Anote a data e o que aconteceu.

Fim da Parte III

Chegamos ao fim da terceira parte. Vimos que OpenAI, Anthropic e Google disputam a mesma fronteira com missões e arquiteturas de poder diferentes. Entendemos que o Google não está parado e que a China não é apenas DeepSeek. Comparamos quatro estratégias nacionais para a mesma revolução. Acompanhamos a tentativa de Musk de juntar todas as camadas num só grupo, a mudança de endereço dos talentos e a corrida por capital. E,

agora, aprendemos com Nokia e BlackBerry que nenhuma liderança está a salvo quando a interface muda.

Entender a disputa era o terceiro passo. O quarto é ver o que ela está fazendo com os negócios.

Porque essa corrida não fica dentro dos laboratórios. Ela chega ao mundo real em forma de velocidade. Empresas novas, com equipes pequenas, estão indo da ideia ao produto, do produto à distribuição global e da distribuição à receita em um tempo que, na minha carreira, levaria anos. Isso não elimina o risco nem garante a sobrevivência. Mas muda o relógio de quem empreende, de quem compete e de quem contrata.

Na Parte IV, vamos olhar para essa compressão dos negócios. Começando por uma pergunta que, poucos anos atrás, pareceria absurda: como uma empresa sai do zero e chega a US$ 100 milhões em receita em questão de meses?

Notas

1. BBC News. "Samsung overtakes Nokia in mobile phone shipments." 27 abr. 2012 (dados da Strategy Analytics: liderança da Nokia desde 1998, quando superou a Motorola; no 1º tri. de 2012, Samsung 93 milhões e Nokia quase 83 milhões de aparelhos). https://www.bbc.com/news/business-17865117; Korea Herald. "Samsung ends Nokia's 14-year phone lead." Abr. 2012. https://www.koreaherald.com/article/10362300

2. Gartner, dados de vendas mundiais de 2009, reproduzidos em: TechCrunch. "Smartphone Sales Up 24 Percent, iPhone's Share Nearly Doubled Last Year (Gartner)." 23 fev. 2010. https://techcrunch.com/2010/02/23/smartphone-iphone-sales-2009-gartner/

3. USA Today. "CEO Forum: Microsoft's Ballmer having a 'great time'." 29 abr. 2007.

https://usatoday30.usatoday.com/money/companies/management/2007-04-29-ballmer-ceo-forum-usat_N.htm

4. Elop, Stephen. Memorando interno da Nokia ("burning platform"), fev. 2011. Texto integral em: The Sydney Morning Herald. "Full text: Nokia CEO Stephen Elop's 'burning platforms' memo." 18 jun. 2015. https://www.smh.com.au/technology/full-text-nokia-ceo-stephen-elops-burning-platforms-memo-20150618-ghqwr0.html

5. BBC News. "Nokia and Microsoft form partnership." 11 fev. 2011. https://www.bbc.co.uk/news/business-12427680

6. Microsoft. "Microsoft to acquire Nokia's devices & services business, license Nokia's patents and mapping services." 3 set. 2013. https://news.microsoft.com/source/2013/09/03/microsoft-to-acquire-nokias-devices-services-business-license-nokias-patents-and-mapping-services/; Nokia. "Nokia completes sale of substantially all of its Devices & Services business to Microsoft." 25 abr. 2014. https://www.nokia.com/newsroom/nokia-completes-sale-of-substantially-all-of-its-devices--services-business-to-microsoft/

7. CNBC. "BlackBerry will stop designing its own phones and focus on software." 28 set. 2016. https://www.cnbc.com/2016/09/28/blackberry-will-stop-designing-its-own-phones-and-focus-on-software.html

8. Google/Motorola Mobility. "Google Acquires Motorola Mobility." PR Newswire, 22 maio 2012. https://www.prnewswire.com/news-releases/google-acquires-motorola-mobility-152530155.html; Lenovo. "Lenovo Completes Acquisition of Motorola Mobility from Google." 30 out. 2014. https://news.lenovo.com/pressroom/press-releases/lenovo-completes-full-acquisition-motorola-mobility-from-google/

9. Nokia. "Nokia Corporation Financial Report for Q4 2025 and full year 2025." 29 jan. 2026 (vendas líquidas de 2025: 19,889 bilhões de euros). https://www.nasdaq.com/press-release/nokia-corporation-financial-report-q4-2025-and-full-year-2025-2026-01-29

10. Nokia. "Inside Information: NVIDIA to make USD 1.0 billion equity investment in Nokia in addition to new strategic partnership." 28 out. 2025. https://www.nokia.com/newsroom/inside-information-nvidia-to-make-

usd-1-billion-equity-investment-in-nokia-in-addition-to-new-strategic-partnership-nokias-board-resolved-on-directed-share-issuance-to-nvidia/

11. BlackBerry. "BlackBerry Reports Fourth Quarter and Full Fiscal Year 2026 Results." 9 abr. 2026 (ano fiscal encerrado em 28 fev. 2026). https://www.nasdaq.com/press-release/blackberry-reports-fourth-quarter-and-full-fiscal-year-2026-results-2026-04-09

12. Apple. "Apple Reinvents the Phone with iPhone." 9 jan. 2007. https://www.apple.com/newsroom/2007/01/09Apple-Reinvents-the-Phone-with-iPhone/

13. Apple. "iPhone 3G on Sale Tomorrow." 10 jul. 2008. https://www.apple.com/newsroom/2008/07/10iPhone-3G-on-Sale-Tomorrow/

14. NTT Docomo. "History." Consultado em 6 out. 2026 (lançamento do i-mode em fev. 1999). https://www.docomo.ne.jp/english/corporate/about/outline/history/index.html

15. Apple. "Apple reports first quarter results." 29 jan. 2026. https://www.apple.com/newsroom/2026/01/apple-reports-first-quarter-results/

16. Apple. "Siri AI, a profoundly more capable and personal assistant powered by the next generation of Apple Intelligence, is here." 14 set. 2026. https://www.apple.com/newsroom/2026/09/siri-ai-a-profoundly-more-capable-and-personal-assistant-is-here/

17. Capoot, Ashley; Rooney, Kate; Eudaily, Chris. "OpenAI DevDay recap: AI lab rolls out Dots agents, Altman and Friar comment on IPO." CNBC, 29–30 set. 2026. https://www.cnbc.com/2026/09/29/openai-devday-2026-live-updates.html

18. Hardwick, Tim. "OpenAI's Jony Ive-Designed Device Delayed to 2027." MacRumors (com base em reportagem da Wired), 10 fev. 2026. https://www.macrumors.com/2026/02/10/openais-jony-ive-designed-device-delayed-to-2027/

19. Wang, Sarah; Kahl, Justin; Xu, Shangda. "Leaders, gainers and unexpected winners in the Enterprise AI arms race." a16z, 30 jan. 2026. https://www.a16z.news/p/leaders-gainers-and-unexpected-winners

PARTE IV

A compressão dos negócios

CAPÍTULO 17

De zero a US$100 milhões em meses

Tese do capítulo: a IA comprime o tempo entre ideia, produto, distribuição global e receita, embora não elimine risco nem garanta sobrevivência.

Terminei a Parte III com uma pergunta que, poucos anos atrás, pareceria absurda: como uma empresa sai do zero e chega a US$ 100 milhões em receita em questão de meses? A resposta curta é que algumas estão chegando lá, ou pelo menos dizendo que chegaram, a um ritmo que eu nunca tinha visto. A resposta longa exige cuidado com cada palavra dessa frase, a começar por "receita".

Ao longo da minha carreira, construí empresas e liderei uma indústria nos Estados Unidos. Quem constrói empresas sabe o que é contratar a primeira equipe, conseguir o primeiro cliente, abrir o primeiro canal de venda. E sabe quanto tempo isso costumava levar.

Em 2006, quando chegamos à Flórida, compramos uma pequena empresa de granito em Port Charlotte, e eu mergulhei em todos os aspectos do negócio. Aquele começo cresceu até virar três fábricas de beneficiamento de granito e, mais tarde, duas empresas de reforma. Chegamos a atender a Home Depot, a Lowe's e mais de 150 empresas de armários e reformas no

sudoeste da Flórida. Tudo isso enquanto atravessávamos a jornada dos vistos de trabalho e, a partir de 2010, sem o meu marido.

Hoje, leio sobre empresas com poucos anos de vida que chegam a centenas de milhões de dólares em receita anualizada com equipes que caberiam num andar de escritório. A primeira reação é de espanto. A segunda, de quem já viu muitos números bonitos, é perguntar: o que exatamente está sendo medido?

1. Lovable, Sierra e Cursor como sinais de uma nova curva de crescimento

Três empresas ajudam a enxergar essa curva, cada uma em um mercado diferente.

A **Lovable**, da Suécia, permite que qualquer pessoa crie aplicativos e sites descrevendo o que quer, em linguagem comum. Lançada no fim de 2024, anunciou em julho de 2025 que tinha passado de US$ 100 milhões de receita recorrente anualizada, o ARR, cerca de oito meses depois do lançamento.[1] Em junho de 2026, disse estar perto de US$ 500 milhões. Em 24 de setembro, segundo um dos fundadores, passou de US$ 600 milhões.[2] A empresa afirma que funcionários de quase dois terços das 500 maiores empresas americanas já a usam e que, desde o lançamento, pessoas criaram mais de 60 milhões de projetos na plataforma.[3] Entre os exemplos que a própria Lovable divulga está o de uma brasileira, Sabrine Matos, que sem formação técnica criou um aplicativo de segurança para mulheres.[4]

A **Sierra**, fundada por Bret Taylor, ex-co-CEO da Salesforce e presidente do conselho da OpenAI, e por Clay Bavor, ex-executivo do Google, vende agentes de IA para o atendimento ao cliente de grandes empresas. Em novembro de 2025, anunciou US$ 100 milhões de ARR sete trimestres depois de lançar o produto, em

fevereiro de 2024, e escreveu que isso aconteceu "bem mais rápido" do que esperava.[5] Em maio de 2026, Taylor informou que a empresa tinha chegado a US$ 200 milhões, apenas dois trimestres depois.[6]

A **Cursor**, ferramenta de programação com IA, informou em junho de 2025 mais de US$ 500 milhões de ARR.[7] Em novembro, disse ter passado de US$ 1 bilhão em receita anualizada, com mais de 300 funcionários.[8] Em junho de 2026, segundo a Forbes, chegou a US$ 4 bilhões. E, como vimos no Capítulo 13, em agosto passou a fazer parte da SpaceX, num negócio de US$ 60 bilhões.[9,10]

Não são casos isolados. Segundo a carta anual da empresa de pagamentos Stripe, o número de empresas novas que atingiram US$ 10 milhões de ARR nos primeiros três meses dobrou de 2024 para 2025. E 20% das empresas abertas pela sua ferramenta de constituição de empresas cobraram o primeiro cliente em até 30 dias, contra 8% em 2020.[11]

O tempo entre ter uma ideia e cobrar por ela nunca foi tão curto.

Por que isso acontece? Em parte, pelo que vimos nas partes anteriores. Um produto de IA nasce global: funciona em vários idiomas, é distribuído pela internet e cobrado por assinatura com cartão de crédito. Não precisa de fábrica, de estoque nem de uma rede de lojas. E boa parte do trabalho pesado, o modelo, é alugada de laboratórios como OpenAI, Anthropic e Google. A empresa constrói a camada que o cliente vê e paga pelo cérebro que está por baixo.

2. ARR não é lucro, caixa nem receita auditada de um ano inteiro

No Capítulo 15, separamos as palavras do dinheiro. Vale relembrar a principal: ARR é uma fotografia multiplicada. Se a empresa fatura hoje US$ 50 milhões por mês em assinaturas, pode dizer que tem US$ 600 milhões de ARR. Isso mede a velocidade do momento. Não diz quanto ela faturou no ano, quanto sobrou depois dos custos, nem se os clientes de hoje estarão lá no ano que vem.

NÃO CONFUNDA

ARR é o ritmo atual da receita recorrente, multiplicado por doze meses: uma fotografia, não um filme.

Receita auditada é o que a empresa de fato faturou num período fechado, conferido por auditores independentes.

Margem bruta é o que sobra da receita depois do custo direto de entregar o produto, como os modelos de IA alugados.

Churn é a taxa de clientes que cancelam. Um ARR alto com churn alto é um balde furado.

Há pelo menos três motivos para ler esses números com calma.

O primeiro é que quase todos são divulgados pelas próprias empresas, que são privadas e não publicam balanços auditados. A Sierra, a Lovable e a Cursor podem estar dizendo a verdade, e não tenho motivo para supor o contrário. Mas o leitor deve saber de onde vem o número. Há exemplos do que pode dar errado. Em 2025, a Builder.ai, uma startup britânica de criação de aplicativos com IA que tinha a Microsoft entre os investidores e já tinha sido avaliada em cerca de US$ 1,5 bilhão, entrou em insolvência. Segundo a Bloomberg, a empresa tinha projetado aos credores

US$ 220 milhões em vendas para 2024. A receita real ficou em torno de US$ 50 milhões.[12]

O segundo motivo é o churn. ARR mede o ritmo de hoje, e esse ritmo pode mudar de uma hora para outra. Em 2025, o próprio presidente da Lovable contou que a empresa tinha "perdido US$ 1,5 milhão de ARR em um único dia" ao mudar os clientes de um plano para outro mais barato.[1] Em produtos que se assinam com um clique, também se cancela com um clique.

O terceiro motivo é a margem. Em agosto de 2025, uma reportagem do TechCrunch ouviu investidores e fundadores que diziam que as margens das ferramentas de programação com IA eram "neutras ou negativas": em alguns casos, custava mais rodar o produto do que a empresa conseguia cobrar por ele, porque cada pedido do cliente consome os modelos mais caros dos laboratórios.[13] Algumas dessas empresas passaram a treinar modelos próprios para reduzir essa dependência. É uma situação que conheço de outro mundo: vender muito não é o mesmo que ganhar dinheiro, quando cada unidade vendida custa quase o que se cobra por ela.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: imagine que você vendeu R$ 500 de brigadeiro no primeiro mês de aula e anunciou que tem "um negócio de R$ 6 mil por ano". Pode ser verdade, se continuar vendendo assim. Mas ainda falta descontar o chocolate, o leite condensado, as forminhas, e torcer para os colegas não enjoarem nas férias.

Para quem tem 65 anos: é como a loja que lota na inauguração. O movimento do primeiro mês é ótimo sinal, mas o dono só sabe se o negócio é bom depois de pagar o aluguel, os fornecedores e os funcionários por um ano inteiro, e ver quantos clientes voltaram.

3. Quando o crescimento serve também para atrair capital, clientes e talentos

Há ainda um ponto mais sutil. Nessas empresas, anunciar o crescimento não é só prestar contas. É parte da estratégia.

Veja a sequência da Lovable. Em julho de 2025, ela levantou US$ 200 milhões com valuation de US$ 1,8 bilhão. Dias depois, anunciou os US$ 100 milhões de ARR.[1,14] Em dezembro, levantou US$ 330 milhões a US$ 6,6 bilhões. Em agosto de 2026, US$ 400 milhões a US$ 13,3 bilhões, com investidores da Europa, da Ásia, dos Estados Unidos e da América Latina, entre eles a Kaszek.[2,3] A Sierra seguiu o mesmo ritmo: valia US$ 10 bilhões em setembro de 2025 e, em maio de 2026, levantou US$ 950 milhões a US$ 15,8 bilhões.[6,15] Quando a Sierra anunciou seus primeiros US$ 100 milhões de ARR, o TechCrunch calculou que ela valia cem vezes essa receita anualizada.[15]

Cada anúncio de crescimento ajuda a trazer o próximo cheque. O cheque permite contratar os melhores engenheiros, pagar os modelos mais caros e fazer mais marketing. Isso gera mais crescimento, que gera o próximo anúncio. Para os clientes corporativos, um número grande também tranquiliza: ninguém quer apostar seu atendimento ou seu código numa ferramenta que pode fechar no ano seguinte. E, para os talentos, como vimos no Capítulo 14, o endereço que cresce mais rápido é o que atrai.

Esse ciclo não é fraude nem mágica. É a forma como o capital de risco sempre funcionou, só que muito mais rápida. Mas ele tem um risco embutido: se o crescimento desacelera, o ciclo pode girar ao contrário. Menos capital, menos contratações, menos produto, menos crescimento.

E há um risco que é próprio da IA. Muitas dessas empresas dependem dos mesmos laboratórios que, às vezes, se tornam seus concorrentes. A OpenAI tem o Codex, e a Anthropic tem o Claude Code, ambos disputando o mesmo programador que usa a Cursor.[13] A Windsurf, rival da Cursor, chegou a negociar a venda para a OpenAI em 2025. O negócio não saiu, e os fundadores e parte da equipe acabaram indo para o Google.[13] Crescer rápido em cima do motor de outra empresa é possível, mas exige saber o que acontece quando o dono do motor resolve entrar no mesmo mercado.

FOTOGRAFIA DA FRONTEIRA — A CURVA COMPRIMIDA (6 DE OUTUBRO DE 2026)

Lovable: mais de US$ 600 milhões de ARR (24 set. 2026), segundo a empresa; valuation de US$ 13,3 bilhões (ago. 2026).

Sierra: US$ 200 milhões de ARR (maio 2026), segundo a empresa; valuation de US$ 15,8 bilhões (maio 2026).

Cursor: US$ 4 bilhões de receita anualizada (jun. 2026, segundo a Forbes); comprada pela SpaceX por US$ 60 bilhões (ago. 2026).

Stripe (2025): o dobro de empresas novas chegando a US$ 10 milhões de ARR em três meses, em relação a 2024.

Fatos vivos: confira o Atlas Vivo da IA.

O QUE AINDA NÃO SABEMOS

Quantos desses clientes ainda estarão pagando daqui a dois anos. Se as margens vão melhorar à medida que o custo dos modelos cair, ou piorar à medida que os clientes pedirem modelos cada vez mais caros. Se os laboratórios de fronteira vão absorver os mercados dessas empresas. E quanto do crescimento é demanda duradoura, e quanto é curiosidade de quem quis experimentar a novidade.

O que mudou / O que significa / O que fazer agora

O que mudou: empresas de IA estão indo da ideia ao produto, do produto ao mundo e do mundo à receita em meses, não em anos. Lovable, Sierra e Cursor mostram essa curva em mercados diferentes, e os dados da Stripe sugerem que ela não se limita a três estrelas.

O que significa: a velocidade de começar mudou de escala, mas as perguntas antigas continuam valendo. Quanto o cliente paga, quanto custa entregar, quanto tempo ele fica e quem pode tomar o seu lugar. ARR é um sinal de velocidade, não de valor construído.

O que fazer agora: quando ler que uma empresa "chegou a US$ 100 milhões em meses", pergunte: é ARR ou receita? Quem divulgou o número? Qual é a margem? Quantos clientes cancelam? E, se você empreende, use essa nova velocidade a seu favor para testar ideias rápido e barato, mas não confunda a primeira onda de vendas com um negócio de pé. A velocidade de começar aumentou. A obrigação de construir valor real não desapareceu.

AÇÃO EM DEZ MINUTOS

Escolha uma ideia de produto ou serviço que você tem guardada há anos. No seu Diário de Fronteira, escreva quanto tempo e quantas pessoas ela teria exigido há dez anos. Depois, peça a uma IA um plano para testá-la em trinta dias, com o menor custo possível. Anote a data, o que mudou no seu cálculo e uma pergunta que você faria antes de chamar o resultado de negócio.

Mas há um caso extremo dessa compressão que merece um capítulo próprio. Se pouco mais de trezentas pessoas puderam levar uma empresa a US$ 1 bilhão de receita anualizada, até onde dá para reduzir essa equipe? No próximo capítulo, vamos falar da empresa de uma pessoa.

Notas

1. Heim, Anna. "Eight months in, Swedish unicorn Lovable crosses the $100M ARR milestone." TechCrunch, 23 jul. 2025. https://techcrunch.com/2025/07/23/eight-months-in-swedish-unicorn-lovable-crosses-the-100m-arr-milestone/; Lovable. "$100M ARR & Lovable Agent." 23 jul. 2025. https://lovable.dev/blog

2. Mehta, Ivan. "Lovable's annualized revenue crosses $600M as vibe coding takes off." TechCrunch, 24 set. 2026. https://techcrunch.com/2026/09/24/lovables-annualized-revenue-crosses-600m-as-vibe-coding-takes-off/

3. Lovable. "We just raised $400M in Series C funding to help people run their businesses." 12 ago. 2026. https://lovable.dev/blog/series-c; Bort, Julie. "Lovable confirms new $13.3B valuation, raises another $400M." TechCrunch, 12 ago. 2026. https://techcrunch.com/2026/08/12/lovable-confirms-new-13-3b-valuation-raises-another-400m/

4. Oliver, Ash. "How Sabrine used Lovable to Fight Gender-Based Violence in Brazil." Lovable, 18 set. 2025. https://lovable.dev/blog

5. Sierra. "Sierra hits $100M ARR milestone in 7 quarters." 21 nov. 2025. https://sierra.ai/blog/100m-arr

6. Taylor, Bret. Publicação no LinkedIn: "Sierra just hit $200M in ARR." Maio 2026. https://www.linkedin.com/posts/brettaylor_sierra-just-hit-200m-in-arr-it-took-seven-activity-7466233808489566209-QsxX; CNBC. "Bret Taylor's Sierra raises nearly $1 billion months after last capital push." 4 maio 2026. https://www.cnbc.com/2026/05/04/bret-taylor-sierra-fundraise-openai.html

7. Cursor. "Series C and Scale." 6 jun. 2025. https://cursor.com/blog/series-c

8. Cursor. "Past, Present, and Future." 13 nov. 2025. https://cursor.com/blog/series-d

9. Nieva, Richard. "Cursor Hits $4 Billion Annualized Revenue Ahead Of SpaceX IPO." Forbes, 8 jun. 2026. https://www.forbes.com/sites/richardnieva/2026/06/08/cursor-4-billion-annualized-revenue/

10. Ha, Anthony. "SpaceX officially closes its Cursor acquisition." TechCrunch, 15 ago. 2026. https://techcrunch.com/2026/08/15/spacex-officially-closes-its-cursor-acquisition/

11. Bort, Julie. "More startups are hitting $10M ARR in 3 months than ever before." TechCrunch, 24 fev. 2026 (sobre a carta anual de 2025 da Stripe). https://techcrunch.com/2026/02/24/more-startups-are-hitting-10m-arr-in-3-months-than-ever-before/; Stripe. "Stripe publishes 2025 annual letter." Fev. 2026. https://stripe.com/newsroom/news/stripe-2025-update

12. Bloomberg, reproduzido por The Straits Times. "Builder.ai, an AI start-up with operations in Singapore, overestimated sales by 300%." 23 maio 2025. https://www.straitstimes.com/business/companies-markets/builder-ai-ai-start-up-with-operations-in-singapore-overestimated-sales-by-300; Silberling, Amanda. "Once worth over $1B, Microsoft-backed Builder.ai is running out of money." TechCrunch, 20 maio 2025. https://techcrunch.com/2025/05/20/once-worth-over-1b-microsoft-backed-builder-ai-is-running-out-of-money/

13. Temkin, Marina. "High costs and thin margins threatening AI coding startups." TechCrunch, 7 ago. 2025. https://techcrunch.com/2025/08/07/the-high-costs-and-thin-margins-threatening-ai-coding-startups/

14. Lovable. "Lovable Raises $200M, Valued at $1.8B, Just Eight Months After Launch." 17 jul. 2025; "Lovable raises $330M to power the age of the builder." 18 dez. 2025. https://lovable.dev/blog

15. Temkin, Marina. "Bret Taylor's Sierra reaches $100M ARR in under two years." TechCrunch, 21 nov. 2025. https://techcrunch.com/2025/11/21/bret-taylors-sierra-reaches-100m-arr-in-under-two-years/

CAPÍTULO 18

A empresa de uma pessoa

Tese do capítulo: uma pessoa apoiada por agentes poderá operar com alcance antes reservado a pequenas organizações, mas continuará responsável por julgamento e consequência.

Terminei o capítulo anterior perguntando até onde dá para reduzir uma equipe. Em 2 de abril de 2026, o New York Times publicou uma resposta que circulou o mundo. Contava a história de Matthew Gallagher, que, da própria casa em Los Angeles, começou uma empresa de telemedicina com US$ 20 mil, dois meses de trabalho e mais de uma dúzia de ferramentas de IA. Segundo o jornal, ele usou IA para escrever o código, os textos do site, as imagens e os vídeos dos anúncios e o atendimento ao cliente. A empresa, a Medvi, vende remédios para emagrecimento da família do Ozempic. Em 2025, seu primeiro ano completo, faturou US$ 401 milhões. Para 2026, estava a caminho de US$ 1,8 bilhão. Com um único funcionário: o irmão.[1]

A história parecia confirmar uma previsão que Sam Altman, da OpenAI, tinha feito em 2024: a de que logo veríamos a primeira empresa de um bilhão de dólares com uma só pessoa.[2] Mas, poucos dias depois, o próprio jornal acrescentou uma nota ao texto. Ele deveria ter informado que a Medvi enfrentava ações regulatórias e judiciais. Em fevereiro, a FDA, a agência americana de medicamentos, tinha enviado à empresa uma carta de advertência por afirmações que considerou falsas ou enganosas

sobre os remédios manipulados oferecidos num site com a marca Medvi. A empresa respondeu que o site era de um afiliado.[1,3,4]

Guardo as duas metades dessa história juntas, porque elas são o assunto deste capítulo. A IA pode dar a uma pessoa o alcance de uma organização. Não lhe dá uma organização para dividir a responsabilidade.

Ao longo da vida, construí empresas e liderei uma indústria nos Estados Unidos. Hoje uso inteligência artificial em decisões, pesquisa, comunicação e criação. E sinto, todos os dias, a distância entre o que eu consigo tentar sozinha e o que eu consigo garantir sozinha.

Na minha vida, a IA não é um assunto que eu estudo de longe. Ela já está dentro do trabalho. Na Freedom to Belong, criamos a entrevista com uma oficial de IA: uma oficial virtual faz até 20 perguntas em voz alta, em inglês, exatamente como na entrevista do USCIS, para que o imigrante possa treinar antes do dia de verdade. No I Need to Call My Mother, a IA organiza as histórias, sugere perguntas, revisa o texto e cria rascunhos de imagem. Mas quem decide é a família.

1. Pesquisa, design, código, marketing, atendimento, finanças e operações como funções ampliadas

Durante a maior parte da minha carreira, cada uma dessas palavras era um departamento. Pesquisa de mercado era uma agência contratada. Design era um estúdio. Código era uma equipe de tecnologia, ou um fornecedor caro. Marketing, atendimento, financeiro e operações eram pessoas, salas e orçamentos.

Hoje, cada uma dessas funções pode ser ampliada por sistemas de IA. Não substituída: ampliada. Uma pessoa pode pedir a um assistente que leia cinquenta relatórios e resuma o que importa. Pode descrever um aplicativo, como vimos com a Lovable no capítulo anterior, e recebê-lo funcionando. Pode gerar a primeira versão de uma campanha, de um contrato, de uma planilha de fluxo de caixa ou de um manual de atendimento. O resultado ainda precisa de revisão, mas a primeira versão deixou de custar semanas.

Os números mostram que mais gente está tentando. Segundo a Carta, plataforma que administra a participação societária de dezenas de milhares de empresas americanas, a fatia de novas startups com um único fundador subiu de 23,7% em 2019 para 36,3% no primeiro semestre de 2025. A própria Carta atribui parte dessa alta à IA, que ampliou o que uma pessoa consegue fazer em tempo limitado. Mas o mesmo relatório traz um contrapeso: em 2024, as empresas de fundador único eram 30% das novas startups e receberam só 14,7% do dinheiro investido em rodadas com preço definido.[5] Começar sozinho ficou mais fácil. Convencer investidores de que dá para crescer sozinho, nem tanto.

E há o caso que muita gente cita como prova. Em junho de 2025, a Wix comprou a Base44, uma plataforma israelense de criação de aplicativos com IA, por cerca de US$ 80 milhões, mais pagamentos condicionados a metas até 2029. O fundador, Maor Shlomo, tinha começado o projeto sozinho, como trabalho paralelo, seis meses antes, sem investidores externos. Segundo a Wix, a empresa tinha oito funcionários quando foi vendida.[6] No primeiro trimestre de 2026, segundo o jornal israelense Globes, a Base44 bateu as metas, e a Wix pagou mais US$ 38 milhões em ações ao fundador e a funcionários.[7]

NÃO CONFUNDA

Fundador único é quem abre a empresa sozinho. Ele pode, e muitas vezes vai, contratar pessoas depois.

Empresa de uma pessoa é a que opera com uma pessoa no centro e o resto do trabalho feito por ferramentas, agentes e parceiros externos.

Delegar uma tarefa é pedir que outro faça. *Delegar a responsabilidade* não existe: diante do cliente, da lei e do banco, ela continua sendo sua.

2. O que pode ser delegado; o que exige responsabilidade humana

No Capítulo 3, sugeri tratar um agente como um funcionário novo e muito rápido: instruções claras, limites definidos, supervisão no começo e confiança construída com o tempo. Os casos reais mostram por que esse cuidado não é exagero.

Em 2025, a Anthropic e a empresa de avaliação Andon Labs colocaram uma versão do Claude para administrar uma pequena loja no escritório da Anthropic em San Francisco, com a missão de dar lucro. Na primeira fase, o lojista artificial, apelidado de Claudius, perdeu dinheiro. Se deixou convencer por funcionários a dar descontos e brindes, vendeu produtos abaixo do custo e chegou a orientar clientes a pagar numa conta que ele mesmo inventou. A própria Anthropic escreveu que, se fosse contratar alguém para a loja naquele momento, não contrataria o Claudius.[8] Na segunda fase, publicada em dezembro de 2025, com modelos mais novos, melhores ferramentas e outros agentes ao redor, as semanas de prejuízo praticamente desapareceram. Mas a loja continuou ingênua. O agente e o "diretor executivo" artificial

criado para supervisioná-lo chegaram a aceitar um contrato de compra futura de cebolas, até que um funcionário avisou que uma lei americana de 1958 proíbe exatamente esse tipo de contrato. A conclusão da Anthropic: a distância entre "capaz" e "completamente robusto" continua grande.[9]

Outro episódio virou alerta para quem cria software sem saber programar. Em julho de 2025, Jason Lemkin, fundador da comunidade de empreendedores SaaStr, contou que o agente da plataforma Replit apagou o banco de dados de produção do projeto dele, mesmo depois de ele ter instruído, segundo ele, onze vezes, em letras maiúsculas, que nada fosse alterado sem permissão. O agente ainda disse, erradamente, que não era possível recuperar os dados.[10] O presidente da Replit, Amjad Masad, pediu desculpas, disse que aquilo era inaceitável e anunciou novas proteções. E Lemkin registrou outro detalhe que interessa a quem faz contas: em três dias e meio, tinha gastado US$ 607,70 além da assinatura mensal de US$ 25.[10]

Há ainda o risco que vem de fora. Em dezembro de 2025, a OpenAI escreveu que a injeção de instruções, a técnica de esconder comandos maliciosos em páginas, documentos ou e-mails para enganar um agente, "dificilmente será um dia totalmente resolvida", assim como os golpes na internet. A recomendação da empresa é dar instruções específicas, evitar acesso amplo e exigir confirmação antes de ações sensíveis, como pagar ou enviar mensagens.[11]

E a produtividade, quando medida com rigor, é menos óbvia do que parece. Num estudo controlado publicado em julho de 2025 pelo METR, organização independente de pesquisa, programadores experientes levaram 19% mais tempo para concluir tarefas quando podiam usar as ferramentas de IA do início de 2025, embora acreditassem estar 20% mais rápidos. Em fevereiro

de 2026, o METR publicou dados novos sugerindo que as ferramentas mais recentes provavelmente já aceleram o trabalho, mas avisou que o próprio método tinha ficado pouco confiável para medir quanto.[12]

A IA pode multiplicar o que você faz. Não pode assinar por você.

O que fica, então, com a pessoa? Na minha experiência, quatro coisas. A **escolha** do que vale a pena fazer. A **verificação** de tudo o que chega ao cliente, ao governo ou ao banco. A **relação** com quem confia em você: clientes, parceiros, funcionários, comunidade. E a **consequência**, porque a carta da FDA tinha o nome da empresa, não o de uma ferramenta.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: imagine que você pode chamar ajudantes incansáveis para o trabalho de grupo: um pesquisa, outro monta os slides, outro revisa o texto. O trabalho fica pronto em uma tarde. Mas, se houver uma informação errada ou copiada, quem apresenta para a turma, e leva a nota, é você.

Para quem tem 65 anos: é como ter um escritório inteiro de contadores, secretárias e vendedores à disposição, que trabalham rápido e nunca dormem. Só que nenhum deles tem carteira assinada nem responde por nada. Quem assina o balanço continua sendo o dono.

3. Do "solopreneur" ao fundador com uma organização artificial ao redor

A palavra em inglês para quem empreende sozinho, *solopreneur*, sugere uma pessoa trabalhando isolada. Os casos reais mostram outra coisa.

A Medvi não fazia tudo sozinha. Segundo as reportagens, médicos e farmácias de plataformas parceiras de telemedicina, como a OpenLoop e a CareValidate, cuidavam das consultas, das receitas, da entrega dos remédios e de parte da conformidade regulatória. A Medvi ficava com a marca, o marketing e a relação com o cliente.[1,2] A Base44 tinha sido criada por uma pessoa, mas precisou da escala e da distribuição de uma empresa grande para dar o passo seguinte. O próprio Shlomo disse que a escala e o volume de que precisavam não eram algo que conseguiriam alcançar crescendo por conta própria.[6] E, segundo a Carta, os fundadores únicos contratam o primeiro funcionário mais cedo do que os times de fundadores: uma mediana de 399 dias depois da abertura da empresa, contra 480.[5]

O que está surgindo, então, não é a empresa sem ninguém. É uma nova forma de organização. No centro, uma ou poucas pessoas com julgamento, relacionamento e responsabilidade. Ao redor, uma camada de agentes que pesquisam, escrevem, programam, atendem e organizam. E, por fora, parceiros humanos que fazem o que exige licença, presença física, confiança ou assinatura: o médico, o contador, o advogado, o fornecedor, o banco.

FOTOGRAFIA DA FRONTEIRA — A EMPRESA ENXUTA (6 DE OUTUBRO DE 2026)

Fundadores únicos (Carta): 36,3% das novas startups americanas no 1º semestre de 2025, contra 23,7% em 2019.

Medvi (NYT, abr. 2026): US$ 401 milhões em vendas em 2025, com dois funcionários e parceiros externos; carta de advertência da FDA em fev. 2026.

Base44: vendida à Wix em jun. 2025 por cerca de US$ 80 milhões; mais US$ 38 milhões em metas pagos no 1º tri. de 2026.

Agentes: injeção de instruções ainda sem solução definitiva, segundo a OpenAI (dez. 2025).

Fatos vivos: confira o Atlas Vivo da IA.

Hoje eu me divido entre a Forever Realty, a Freedom to Belong, a US Legal Now, da qual sou fundadora, sócia e conselheira, a American Photonics e o I Need to Call My Mother. Para dar conta de tudo, montei várias equipes de agentes de inteligência artificial: um escritor que trabalha comigo nos livros, um que cuida dos sites, um chefe de gabinete, um para o imobiliário, um para os e-mails e um para os investimentos. Uso as IAs americanas e muitas chinesas. E, nos últimos anos, contei também com um assistente executivo, o Daniel, que me ajudou a organizar os projetos que pareciam se multiplicar da noite para o dia.

O QUE AINDA NÃO SABEMOS

Se empresas de uma pessoa vão durar ou se vão precisar contratar quando crescerem. Se os agentes vão ficar confiáveis o bastante para lidar sozinhos com dinheiro, dados de clientes e decisões reguladas. Quanto vai custar, de verdade, manter uma organização artificial funcionando quando o uso aumentar. E como governos vão fiscalizar empresas enormes com quase ninguém dentro.

O que mudou / O que significa / O que fazer agora

O que mudou: pesquisa, design, código, marketing, atendimento, finanças e operações deixaram de exigir, cada uma, uma equipe para começar. Uma pessoa com agentes pode tentar o que antes exigia uma pequena empresa, e mais gente está fazendo isso: mais de um terço das novas startups americanas já nasce com um único fundador.

O que significa: a IA reduz o custo de tentar, não o peso de responder. Os agentes erram, podem ser manipulados e às vezes custam mais do que parecem. Quem está no centro continua responsável por escolher, verificar, cuidar das relações e responder pelas consequências.

O que fazer agora: antes de entregar uma função inteira a um agente, comece por uma tarefa, com limites claros, permissões mínimas e revisão sua antes de qualquer coisa chegar ao cliente. Escreva quem são os seus parceiros humanos indispensáveis e não abra mão deles. A empresa de uma pessoa não significa uma pessoa sem parceiros; significa uma nova proporção entre ambição e estrutura.

AÇÃO EM DEZ MINUTOS

No seu Diário de Fronteira, desenhe duas colunas para um negócio que você tem ou sonha ter. Na primeira, liste as tarefas que um agente de IA poderia fazer ou preparar: pesquisa, rascunhos, planilhas, respostas a clientes. Na segunda, as decisões que só você pode tomar e assinar. Ao lado de cada tarefa da primeira coluna, anote quem revisa e quando. Coloque a data.

Mas a empresa de uma pessoa só é possível por um motivo que ainda não examinamos de frente: a inteligência ficou barata, e continua barateando. No próximo capítulo, vamos entender o que acontece quando o preço da inteligência cai, e quem ganha com isso.

Notas

1. Griffith, Erin. "How A.I. Helped One Man (and His Brother) Build a $1.8 Billion Company." The New York Times, 2 abr. 2026, atualizado em 9 abr. 2026 (com nota sobre a carta da FDA e uma ação coletiva pendente). https://www.nytimes.com/2026/04/02/technology/ai-billion-dollar-company-medvi.html

2. Majić, Josipa. "AI And $20,000 Helped One Man Build A $1.8 Billion Telehealth Startup." Forbes, 2 abr. 2026 (sobre a Medvi, a parceria com OpenLoop e CareValidate e a previsão de Sam Altman, em 2024, de uma empresa de um bilhão de dólares com uma só pessoa). https://www.forbes.com/sites/josipamajic/2026/04/02/ai-and-20000-helped-one-man-build-a-18-billion-telehealth-startup/

3. U.S. Food and Drug Administration. Warning Letter, "MEDVi, LLC dba MEDVi" (MARCS-CMS 721455). 20 fev. 2026. https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/warning-letters/medvi-llc-dba-medvi-721455-02202026

4. Futurism. "AI-Powered Drug Marketer Medvi Responds After Allegations About Fake Doctors and Patients." Abr. 2026.

https://futurism.com/artificial-intelligence/ai-drug-marketer-medvi-responds; Drug Discovery & Development. "The New York Times spotlighted MEDVi. The FDA had already warned the self-proclaimed 'fastest growing company in history.'" Abr. 2026. https://www.drugdiscoverytrends.com/the-new-york-times-spotlighted-medvi-the-fda-had-already-warned-the-self-proclaimed-fastest-growing-company-in-history/

5. Walker, Peter; Shad, Hamza. "Solo Founders Report 2025." Carta, 9 dez. 2025. https://carta.com/data/solo-founders-report/

6. Wix.com. "Wix Further Expands into Vibe Coding with Acquisition of Base44, a Hyper-Growth Startup that Simplifies Web and App Creation with AI." Wix Press Room, 18 jun. 2025. https://www.wix.com/press-room/home/post/wix-further-expands-into-vibe-coding-with-acquisition-of-base44-a-hyper-growth-startup-that-simplif; Bort, Julie. "6-month-old, solo-owned vibe coder Base44 sells to Wix for $80M cash." TechCrunch, 18 jun. 2025. https://techcrunch.com/2025/06/18/6-month-old-solo-owned-vibe-coder-base44-sells-to-wix-for-80m-cash/; Calcalist/Ctech. "Vibe coding fever: Solo entrepreneur's Base44 acquired by Wix for $80 million." Jun. 2025. https://www.calcalistech.com/ctechnews/article/s1iflnlelx

7. Globes. "Wix pays $38m earnout payments for Base44." 2026. https://en.globes.co.il/en/article-wix-pays-38m-earnout-payments-for-base44-1001542919

8. Anthropic. "Project Vend: Can Claude run a small shop? (And why does that matter?)" 27 jun. 2025. https://www.anthropic.com/research/project-vend-1

9. Anthropic. "Project Vend: Phase two." 18 dez. 2025. https://www.anthropic.com/research/project-vend-2

10. Sharwood, Simon. "Vibe coding service Replit deleted user's production database, faked data, told fibs galore." The Register, 21 jul. 2025. https://www.theregister.com/software/2025/07/21/vibe-coding-service-replit-deleted-production-database/719783; Business Insider. "Replit CEO Apologizes After AI Coding Tool Wipes Company's Database." Jul. 2025. https://www.businessinsider.com/replit-ceo-apologizes-ai-coding-tool-delete-company-database-2025-7

11. OpenAI. "Continuously hardening ChatGPT Atlas against prompt injection attacks." 22 dez. 2025. https://openai.com/index/hardening-atlas-against-prompt-injection/; Fortune. "OpenAI says prompt injections that can trick AI browsers like ChatGPT Atlas may never be fully 'solved'." 23 dez. 2025. https://fortune.com/2025/12/23/openai-ai-browser-prompt-injections-cybersecurity-hackers/

12. METR. "Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity." 10 jul. 2025. https://metr.org/blog/2025-07-10-early-2025-ai-experienced-os-dev-study/; METR. "We are Changing our Developer Productivity Experiment Design." 24 fev. 2026. https://metr.org/blog/2026-02-24-uplift-update/

CAPÍTULO 19

Quando o preço da inteligência cai

Tese do capítulo: o custo para construir a fronteira sobe, enquanto o preço para consumir uma capacidade equivalente cai rapidamente.

Terminei o capítulo anterior dizendo que a empresa de uma pessoa só é possível porque a inteligência ficou barata. Agora quero mostrar o tamanho dessa queda com um único exemplo.

Em 31 de janeiro de 2025, a OpenAI lançou o o3, um dos seus primeiros modelos de "raciocínio". Segundo a Epoch AI, instituto independente que acompanha a economia da IA, ele conseguia acertar 75% de uma prova de ciências em nível de doutorado, a GPQA Diamond, gastando em média 30 centavos de dólar por questão. Pouco menos de 18 meses depois, o GPT-5.6 Luna, o mesmo modelo cujo preço vimos cair 80% no Capítulo 2, alcançou o mesmo resultado por quatro centésimos de centavo por questão. Uma queda de 725 vezes. Os próprios pesquisadores compararam: é como se o preço de um carro novo caísse de 50 mil dólares para 69.[1]

Agora guarde esse exemplo ao lado de outro número. Em 2022, o Google gastava 31 bilhões de dólares por ano em investimentos de capital, sobretudo em infraestrutura. Para 2026, a empresa espera gastar cerca de seis vezes isso: entre 180 e 190 bilhões.[2]

As duas coisas são verdade ao mesmo tempo, e é esse o paradoxo deste capítulo. Construir a fronteira nunca foi tão caro. Usar o que a fronteira já alcançou nunca foi tão barato.

Esse movimento me é familiar. Passei boa parte da minha carreira levando tecnologia cara para o consumo de massa: computadores para os supermercados, a linha Aptiva na IBM Brasil, celulares Nokia e Motorola em mais de 1.500 pontos de venda. Vi aparelhos e serviços que pareciam privilégio virarem coisa comum.

1. Bilhões em chips, energia, data centers, dados e pesquisadores

Comecemos pelo lado caro. Treinar um modelo de fronteira exige cinco ingredientes, e todos ficaram mais caros.

Chips e data centers. Segundo a Epoch AI, o custo para treinar os modelos de linguagem mais avançados vem crescendo cerca de 3,5 vezes por ano desde 2020.[3] Em 2024, um estudo do mesmo grupo com pesquisadores de Stanford estimou que o GPT-4, de 2023, tinha custado cerca de 40 milhões de dólares só em equipamento e energia para o treinamento final. Se a tendência continuasse, previam, os maiores treinamentos passariam de 1 bilhão de dólares em 2027.[4] No Capítulo 12, vimos que as grandes plataformas projetavam mais de 735 bilhões de dólares em investimentos para 2026. O Google sozinho responde por uma parte grande disso.

Energia. O AI Index 2026, da Universidade Stanford, calculou que a capacidade elétrica dos data centers de IA chegou a 29,6 gigawatts em 2025, comparável ao pico de consumo do estado de Nova York.[5] A Epoch observa que os maiores data centers de IA já

se aproximam de um gigawatt cada, o suficiente para abastecer cerca de um milhão de casas americanas.[3]

Pesquisadores. No estudo de 2024, os salários das equipes de pesquisa representavam de 29% a 49% do custo de desenvolver modelos como o GPT-4 e o Gemini Ultra.[4] E vimos no Capítulo 14 o que aconteceu depois: talento em IA virou objeto de leilão.

Dados. O material de treinamento também tem preço, às vezes cobrado na Justiça. Em setembro de 2025, a Anthropic fechou um acordo de 1,5 bilhão de dólares com autores que a acusavam de usar livros pirateados, cerca de 3 mil dólares por obra. A aprovação final veio em julho de 2026.[6]

Some tudo isso e fica claro por que a fronteira se concentra em poucas mãos, como antecipei no Capítulo 1. E por que, como vimos no Capítulo 15, empresas que faturam bilhões ainda podem perder outros bilhões.

2. A queda do custo por tarefa, token e nível de capacidade

Agora o lado barato. No Capítulo 2, registrei a queda de mais de 280 vezes medida pelo AI Index 2025 no custo de usar um modelo com capacidade parecida à do GPT-3.5. Não era um ponto fora da curva. Era o começo de uma tendência que os pesquisadores vêm medindo de jeitos diferentes.

Para entender os números, é preciso separar três medidas que costumam ser misturadas.

NÃO CONFUNDA

Preço por token é o que a empresa cobra por cada pedaço de texto processado. É o número da tabela de preços.

Custo por tarefa é quanto você realmente gasta para resolver um problema. Modelos de raciocínio podem cobrar menos por token e usar muito mais tokens.

Custo por nível de capacidade é quanto custa alcançar um determinado desempenho, como acertar 75% de uma prova. É a medida que mostra a queda mais impressionante.

Pela medida mais rigorosa, o custo por tarefa para um mesmo nível de capacidade, a Epoch calculou em setembro de 2026 que o preço caiu cerca de 47% por trimestre desde 2023, ou 13 vezes por ano. Segundo o instituto, nenhuma outra tecnologia transformadora barateou tão rápido: a queda foi seis vezes mais rápida que a da computação e 54 vezes mais rápida que a da eletricidade nos Estados Unidos entre 1892 e 1973.[1] Um estudo do MIT, com outro método, encontrou quedas de 5 a 10 vezes por ano.[7] Os números variam, mas apontam todos na mesma direção.

De onde vem essa queda? Em parte, do hardware. O AI Index 2025 registrou que o custo do hardware de IA caía cerca de 30% ao ano, enquanto a eficiência energética melhorava cerca de 40% ao ano.[8] Em parte, de algoritmos melhores: os pesquisadores do MIT estimaram que só a eficiência dos algoritmos explica uma melhora de cerca de 3 vezes por ano. E em parte, da competição entre empresas.[7]

Mas há uma pegadinha que todo empresário precisa conhecer. O mesmo estudo do MIT mostrou que, enquanto uma capacidade fixa fica mais barata, o custo de usar o modelo mais avançado de cada momento subiu de 3 a 18 vezes por ano. Os modelos de ponta

pensam mais, usam mais tokens e custam mais por tarefa.[7] A inteligência de ontem fica barata. A de hoje continua cara.

E há uma segunda consequência, que os economistas já conhecem desde o carvão do século XIX: quando um recurso fica mais barato, o consumo total costuma crescer. Em maio de 2026, o Google informou que processava mais de 3,2 quatrilhões de tokens por mês, contra cerca de 480 trilhões um ano antes e 9,7 trilhões dois anos antes. Na mesma ocasião, o presidente da empresa, Sundar Pichai, disse que muitas empresas já estavam estourando os orçamentos anuais de tokens, e o ano estava só em maio.[2]

O preço de pensar caiu. A conta de quem pensa muito pode subir.

FOTOGRAFIA DA FRONTEIRA — O PREÇO DA INTELIGÊNCIA (6 DE OUTUBRO DE 2026)

Usar: custo de um mesmo desempenho caindo cerca de 13 vezes por ano desde 2023 (Epoch AI, set. 2026).

Exemplo: 75% na GPQA Diamond por US$ 0,30 por questão (o3, jan. 2025) e por US$ 0,0004 (GPT-5.6 Luna, 2026).

Construir: custo de treinar modelos de fronteira crescendo cerca de 3,5 vezes por ano desde 2020 (Epoch AI).

Investir: Google prevê entre US$ 180 e 190 bilhões em investimentos de capital em 2026, contra US$ 31 bilhões em 2022.

Consumir: Google com mais de 3,2 quatrilhões de tokens por mês (maio de 2026).

Fatos vivos: confira o Atlas Vivo da IA.

3. Do minuto de celular ao minuto de raciocínio: uma analogia imperfeita, mas poderosa

Quem viveu o Brasil dos anos 1990 lembra que telefone era patrimônio. Em março de 1997, uma linha fixa convencional custava 1.117 reais, segundo a Folha de S.Paulo, e havia quem a declarasse no Imposto de Renda e a alugasse como investimento.[9] Em 1998, ano da privatização do sistema Telebrás, o país tinha 17 milhões de linhas fixas e 4,6 milhões de celulares. Vinte anos depois, eram 235,5 milhões de celulares, mais aparelhos do que brasileiros. E o preço do minuto de ligação de celular, segundo o sindicato das operadoras, tinha caído de 1,05 real em 1998 para 9 centavos em 2017, com impostos, sem nem descontar a inflação.[10]

Quando o minuto ficou barato, ninguém passou a falar só um pouco mais. O telefone deixou de ser um recurso racionado e virou infraestrutura. Em cima dele nasceram o comércio pelo celular, os aplicativos, os bancos digitais, o pequeno vendedor que atende pelo WhatsApp.

Acredito que estamos começando a ver o mesmo com o "minuto de raciocínio". Uma análise de contrato, uma pesquisa de mercado, a tradução de um manual, um plano de negócios: tarefas que antes exigiam contratar um especialista e esperar dias agora custam centavos e minutos, pelo menos na primeira versão.

Mas a analogia é imperfeita, e é importante dizer onde. O minuto de ligação é sempre o mesmo produto. O "minuto de raciocínio" não é: a própria Epoch avisa que o preço da inteligência é um conceito mais ambíguo do que o preço da eletricidade, porque cada tarefa é diferente.[1] Uma resposta barata pode estar errada, e corrigi-la custa caro. E a queda do celular

dependeu de competição, regulação e investimento público e privado. A da IA também depende, e nada garante que vá continuar no mesmo ritmo.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: seus avós talvez tenham esperado anos por uma linha de telefone e falado rápido no celular porque cada minuto pesava no bolso. Hoje você manda áudio de cinco minutos sem pensar. Com a IA está acontecendo algo parecido: pedir ajuda a um "especialista" artificial está ficando quase tão barato quanto mandar uma mensagem.

Para quem tem 65 anos: lembra quando um interurbano era coisa para fim de noite ou fim de semana, para pagar menos? A inteligência artificial está passando por uma queda de preço parecida, só que muito mais rápida. O que hoje parece caro e sofisticado pode, em pouco tempo, estar ao alcance de qualquer pequeno negócio.

O QUE AINDA NÃO SABEMOS

Se a queda de preço vai continuar no ritmo atual ou desacelerar, à medida que chips, energia e data centers ficam mais disputados. Se os investimentos gigantes em infraestrutura vão se pagar, ou se parte deles vai virar capacidade ociosa. Se os modelos de ponta vão ficar acessíveis ou se vão se tornar um serviço de luxo para poucos. E quanto da queda medida em provas se traduz em economia no trabalho real.

O que mudou / O que significa / O que fazer agora

O que mudou: a inteligência artificial ficou muito mais barata para usar e muito mais cara para construir. O custo de alcançar

um mesmo desempenho cai cerca de 13 vezes por ano, segundo a Epoch, enquanto o custo de treinar a fronteira sobe cerca de 3,5 vezes por ano e os investimentos em infraestrutura chegam a centenas de bilhões.

O que significa: poucas organizações vão construir os modelos mais avançados, mas muitas vão poder usar capacidades que até pouco tempo atrás eram exclusivas. A vantagem deixa de estar em ter acesso à inteligência e passa a estar em saber usá-la, escolher o modelo certo para cada tarefa e controlar o consumo.

O que fazer agora: não decida com base no preço do ano passado. Revise periodicamente quanto custa a IA que você usa, teste modelos mais baratos para tarefas simples e reserve os mais caros para o que realmente exige. E acompanhe o gasto total, não só o preço por uso. Quando inteligência vira utilidade acessível, pequenas empresas ganham ferramentas que pertenciam a instituições.

AÇÃO EM DEZ MINUTOS

No seu Diário de Fronteira, escolha uma tarefa do seu trabalho que você deixava de fazer por ser cara: uma pesquisa, uma tradução, uma análise. Anote quanto ela custaria contratando alguém e quanto tempo levaria. Depois, faça uma primeira versão com uma ferramenta de IA e anote o custo e o tempo. Coloque a data e repita daqui a seis meses.

Mas, se a inteligência está ficando acessível para todos, surge uma pergunta mais delicada: o que acontece com quem não aprender a usá-la? No próximo capítulo, vamos falar de competição, e de quem saberá competir.

Notas

1. Emberson, Luke; Roodman, David. "The plunging price of thought." Epoch AI, 22 set. 2026 (atualizado em 23 set. 2026). https://epoch.ai/publications/the-plunging-price-of-thought

2. Pichai, Sundar. "I/O 2026: Welcome to the agentic Gemini era." Google (blog), 19 maio 2026. https://blog.google/innovation-and-ai/sundar-pichai-io-2026/

3. Epoch AI. "Trends in AI." Consultado em 6 out. 2026. https://epoch.ai/trends

4. Cottier, Ben et al. "The rising costs of training frontier AI models." arXiv:2405.21015, 2024. https://arxiv.org/abs/2405.21015; Epoch AI. "How much does it cost to train frontier AI models?" 2024. https://epoch.ai/publications/how-much-does-it-cost-to-train-frontier-ai-models

5. Stanford HAI. "The 2026 AI Index Report", capítulo "Research and Development." 2026. https://hai.stanford.edu/ai-index/2026-ai-index-report/research-and-development

6. Associated Press. "Judge approves $1.5 billion copyright settlement between AI company Anthropic and authors." Set. 2025. https://apnews.com/article/anthropic-authors-copyright-judge-artificial-intelligence-9643064e847a5e88ef6ee8b620b3a44c; Reuters. "US judge approves Anthropic's $1.5 billion settlement of copyright lawsuit." 20 jul. 2026. https://www.reuters.com/world/us-judge-approves-anthropics-15-billion-settlement-copyright-lawsuit-2026-07-20/

7. Gundlach, Hans; Lynch, Jayson; Mertens, Matthias; Thompson, Neil (MIT FutureTech). "The Price of Progress: Price Performance and the Future of AI." arXiv:2511.23455. https://arxiv.org/abs/2511.23455

8. Stanford HAI. "The 2025 AI Index Report." Abr. 2025. https://hai.stanford.edu/ai-index/2025-ai-index-report

9. Folha de S.Paulo/Agência JB. "Linhas telefônicas atuais serão valorizadas, diz Motta." 7 mar. 1997. https://www1.folha.uol.com.br/fol/eco/ec07032.htm; Gazeta do Povo.

"Privatização da telefonia: quem perdeu foi a viúva" (artigo de opinião). https://www.gazetadopovo.com.br/opiniao/artigos/privatizacao-da-telefonia-quem-perdeu-foi-a-viuva-enj7rjb4j56oimstjq1hcowaa/

10. Warth, Anne. "Há 20 anos, um telefone chegava a custar US$ 5 mil; veja como tudo mudou." Estadão Conteúdo/UOL Tilt, 30 jul. 2018 (dados do SindiTelebrasil). https://www.uol.com.br/tilt/noticias/redacao/2018/07/30/apos-20-anos-de-privatizacao-uso-do-telefone-mudou.htm

CAPÍTULO 20

Quem souber competir, competirá

Tese do capítulo: o risco não é a pessoa perder valor humano; é seu método de trabalho deixar de produzir valor econômico em comparação com métodos ampliados por IA.

Terminei o capítulo anterior com uma pergunta delicada: se a inteligência está ficando acessível para todos, o que acontece com quem não aprender a usá-la?

Há uma frase em inglês que formulei para responder a essa pergunta: *those who can compete, compete*. Quem pode competir, compete. Na primeira versão, ela soava dura, quase uma sentença. Parecia dizer que alguns vão ganhar e outros vão ficar para trás, e que é assim mesmo.

Com o tempo, percebi que a frase precisava ser refinada, porque a primeira versão confundia duas coisas diferentes: o valor de uma pessoa e o valor econômico do método que ela usa para trabalhar. A pessoa não perde valor. O método pode perder. E método se aprende.

Por isso o título deste capítulo não é "quem pode competir", mas "quem souber competir". Saber pode ser ensinado. E, se pode ser ensinado, a pergunta deixa de ser só individual e passa a ser também social: quem vai ter a chance de aprender?

1. Valor humano não é relevância econômica

Vi essa diferença de perto mais de uma vez. Trabalhei com Nokia, Motorola e BlackBerry, e vi, como contei no Capítulo 16, empresas que dominavam o mercado perderem a liderança em poucos anos quando a interface mudou. As pessoas que trabalhavam ali não ficaram menos inteligentes, menos dedicadas ou menos humanas de um dia para o outro. O que perdeu relevância foi um jeito de fazer telefone, de vender telefone e de pensar o telefone.

Aprendi isso de um jeito curioso, em Monte Carlo. Steve e eu fomos jogar pôquer no cassino dos filmes de James Bond. Nos vestimos como para uma estreia de cinema, prontos para o pôquer mais elegante e mais perigoso da nossa vida. Encontramos uma mesa de oito em que só nós dois sabíamos jogar de verdade. Tivemos uma noite lucrativa e saímos de lá decepcionados. Nada naquela mesa fez uma pergunta difícil ao meu jogo. Foi ali que entendi: jogos fáceis dão dinheiro, mas também enferrujam. Receita confortável e pôquer confortável têm o mesmo efeito colateral, e ele não aparece no resultado da noite. Aparece anos depois, na primeira vez em que você se senta diante de alguém que treinou enquanto você era pago para ficar parado. Desde então, essa é uma regra minha, na mesa e nos negócios: aproveite os jogos fáceis, mas marque de propósito os difíceis.

Com a IA, essa distinção chega a quase todas as profissões ao mesmo tempo. E as empresas já começaram a medir o trabalho por ela. Em abril de 2025, Tobi Lütke, presidente da Shopify, plataforma canadense de comércio eletrônico, divulgou um memorando interno dizendo que o uso da IA passava a ser uma expectativa básica para todos na empresa. Mais do que isso: antes de pedir mais pessoas ou mais recursos, cada equipe teria de mostrar por que a tarefa não podia ser feita com IA.[1]

Repare no que essa regra faz. Ela não diz que as pessoas valem menos. Ela compara métodos. A pergunta passa a ser: este trabalho, feito deste jeito, ainda é a melhor forma de produzir este resultado? É uma pergunta econômica, não moral. Mas quem não souber respondê-la pode sentir as consequências como se fosse pessoal.

NÃO CONFUNDA

Valor humano é o que uma pessoa é: dignidade, caráter, afeto, história. Não depende de produtividade e não deveria ser medido por ela.

Relevância econômica é o que o mercado está disposto a pagar por um trabalho num determinado momento. Ela muda quando surgem métodos mais baratos, rápidos ou bons.

Perder a segunda não significa perder a primeira. Mas, num mundo que paga pela segunda, é preciso cuidar dela.

2. Pessoas que dirigem IA, pessoas dirigidas por IA e pessoas excluídas da IA

Quando penso em como a IA está entrando no trabalho, vejo três grupos. Há quem aprenda a dirigir a inteligência artificial. Há quem seja dirigido por sistemas que não compreende. E há quem fique excluído por renda, idioma, medo, educação ou falta de infraestrutura.

Quem é dirigido pela IA nem sempre perde. Um dos estudos mais citados sobre o tema acompanhou mais de 5 mil atendentes de suporte ao cliente de uma empresa de software. Com um assistente de IA sugerindo respostas, a produtividade, medida em problemas resolvidos por hora, subiu 15% em média,

segundo a versão publicada em 2025 no Quarterly Journal of Economics. Entre os atendentes menos experientes, o ganho foi de cerca de 30%. Os pesquisadores sugerem que a IA espalhou as boas práticas dos melhores atendentes para os novatos. Já os mais experientes tiveram pequenos ganhos de velocidade e pequenas quedas de qualidade.[2] A IA pode ser uma escada para quem está começando. Mas uma escada que outra pessoa desenhou.

Quem está começando também é quem mais sente o aperto. Em agosto de 2026, os economistas Erik Brynjolfsson, Bharat Chandar e Ruyu Chen, de Stanford, atualizaram um estudo com dados de folha de pagamento de milhões de trabalhadores americanos. Não encontraram sinal de desemprego generalizado causado pela IA. Mas o emprego de jovens de 22 a 25 anos nas profissões mais expostas à IA estava cerca de 19% abaixo do que estaria se tivesse acompanhado o de jovens em profissões menos expostas. Em 2025, essa diferença era de 15%. A queda veio sobretudo de menos contratações, não de demissões, e se concentrou nas funções em que a IA substitui tarefas. Onde ela complementa o trabalho, o emprego ficou estável ou cresceu, principalmente entre os mais experientes. Os próprios autores tratam esses dados como sinais de alerta iniciais, os "canários na mina de carvão", e não como prova de causa.[3] No Capítulo 14, vimos o mesmo degrau quebrado nas grandes empresas de tecnologia. Na próxima parte do livro, vamos olhar para ele com calma.

E há quem fique de fora. Segundo a Microsoft, em junho de 2026, 18,8% da população mundial em idade de trabalhar tinha usado algum produto de IA generativa. No Norte Global, eram 28,8%; no Sul Global, 16,2%. E a distância aumentou ao longo do último ano, porque o Norte cresce mais rápido.[4] No Brasil, a pesquisa TIC Domicílios 2025, do Cetic.br, mostrou que 32% dos

usuários de internet, cerca de 50 milhões de pessoas, tinham usado IA generativa. Na classe A, eram 69%; nas classes D e E, 16%. Entre quem tem ensino superior, 59%; entre quem tem apenas o ensino fundamental, 17%. E, entre os que não usavam IA e tinham apenas o fundamental, 65% apontaram a falta de habilidade como motivo.[5]

Valor humano não se mede pelo método. Mas oportunidade, sim.

Esses números me fazem refinar a frase mais uma vez. "Quem souber competir, competirá" é verdade. Mas uma tese responsável precisa vir acompanhada de outra: é tarefa de empresas, escolas, governos e de cada um de nós ampliar o número de pessoas que sabem.

Na Freedom to Belong, é isso que tentamos fazer todos os dias. Oferecemos aulas gratuitas pelo Zoom, livros em inglês e português com voz americana e uma entrevista simulada com uma oficial de IA, para que o imigrante treine antes do dia de verdade. Também escrevi livros sobre a cidadania americana, que estão na Amazon: *Cidadania Americana* e *Cidadania Americana: Só 20 Perguntas*.

3. A nova vantagem: experiência, julgamento e inteligência artificial operando juntas

Se a IA espalha a capacidade de fazer, onde fica a vantagem? Minha resposta: na combinação entre o que a máquina faz bem e o que só a experiência ensina.

Um experimento com 758 consultores do Boston Consulting Group, conduzido por pesquisadores de Harvard, do MIT e de

outras instituições e publicado como estudo de trabalho em 2023, mostrou os dois lados dessa combinação. Em 18 tarefas que estavam dentro da capacidade da IA da época, o GPT-4, os consultores que usaram a ferramenta completaram 12,2% mais tarefas, terminaram 25,1% mais rápido e entregaram resultados com qualidade mais de 40% superior. Mas, numa tarefa escolhida por estar fora dessa capacidade, quem usou IA teve 19 pontos percentuais a menos de chance de chegar à resposta certa do que quem trabalhou sozinho.[6] Os pesquisadores chamaram isso de "fronteira tecnológica irregular": a IA é excelente em algumas tarefas e falha em outras que parecem igualmente difíceis.

Saber de que lado da fronteira está cada tarefa é exatamente o tipo de julgamento que vem da experiência. O estudo de Stanford aponta na mesma direção. O emprego caiu entre jovens em profissões que dependem de conhecimento codificado, aquele que está em livros, manuais e procedimentos. E cresceu entre trabalhadores experientes em profissões que dependem de conhecimento tácito, aprendido na prática, com mentores e com situações reais repetidas.[3]

É por isso que não vejo a experiência como algo que a IA torna obsoleto. Vejo como aquilo que torna a IA útil. Uma pessoa com trinta anos de vendas, de gestão ou de atendimento sabe reconhecer uma resposta que parece boa, mas não funciona com um cliente real. Quem tem essa experiência e aprende a usar a IA ganha alcance. Quem tem a ferramenta sem a experiência precisa de alguém que lhe ensine o que conferir.

E aprender a usar já está deixando de ser opcional. Na União Europeia, o AI Act passou a exigir, desde 2 de fevereiro de 2025, que empresas que desenvolvem ou usam sistemas de IA cuidassem da alfabetização em IA de suas equipes. Em julho de 2026, o Digital Omnibus reescreveu essa regra: ela deixou de exigir um

nível "suficiente" e passou a pedir medidas para apoiar o desenvolvimento dessa alfabetização, com atenção especial às pequenas e médias empresas.[7] A regra ficou mais leve. A direção continua a mesma.

FOTOGRAFIA DA FRONTEIRA — COMPETIR COM IA (6 DE OUTUBRO DE 2026)

Início de carreira (Stanford, dados até jun. 2026): emprego de jovens de 22 a 25 anos em profissões muito expostas à IA cerca de 19% abaixo do esperado; era 15% em 2025.

Uso mundial (Microsoft, jun. 2026): 18,8% da população em idade de trabalhar; 28,8% no Norte Global e 16,2% no Sul Global.

Brasil (TIC Domicílios 2025): 32% dos usuários de internet usaram IA generativa; 69% na classe A e 16% nas classes D e E.

Regra: alfabetização em IA no AI Act europeu, reescrita pelo Digital Omnibus em vigor desde 27 jul. 2026.

Fatos vivos: confira o Atlas Vivo da IA.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: não é a IA que vai tirar o seu lugar. É alguém que sabe usá-la tão bem quanto você e que, além disso, aprendeu a conferir o que ela faz. Use a IA para aprender mais rápido, não para deixar de aprender.

Para quem tem 65 anos: a sua experiência não perdeu valor. Ela é justamente o que permite saber se a resposta da máquina serve ou não. O que falta, muitas vezes, é só a ponte: aprender a usar a ferramenta para que essa experiência chegue mais longe.

O QUE AINDA NÃO SABEMOS

Se a queda nas contratações de jovens é causada pela IA ou por outros fatores que coincidiram com ela. Se as empresas vão recriar o degrau de entrada para formar os profissionais experientes de amanhã. Se a distância entre Norte e Sul Global, e entre ricos e pobres dentro de cada país, vai diminuir ou aumentar. E quanto tempo a experiência humana vai continuar sendo a melhor forma de saber onde a IA erra.

O que mudou / O que significa / O que fazer agora

O que mudou: a IA passou a ser comparada com métodos de trabalho, e as empresas começaram a perguntar se uma tarefa pode ser feita com ela antes de contratar. Os primeiros sinais aparecem no início de carreira e na diferença de acesso entre países, classes e níveis de escolaridade.

O que significa: o risco não é perder valor humano, e sim depender de um método que deixou de ser competitivo. A vantagem passa a estar na combinação entre experiência, julgamento e IA. E a tese "quem souber competir, competirá" só é justa se mais pessoas tiverem a chance de saber.

O que fazer agora: identifique no seu trabalho uma tarefa feita do jeito antigo e teste como ela seria com IA. Depois, pergunte onde a sua experiência faz diferença no resultado e invista nela. E, se você tem uma equipe, uma família ou uma comunidade, ajude pelo menos uma pessoa a dar o primeiro passo. Quem souber ampliar sua inteligência competirá em uma escala antes reservada às grandes instituições.

AÇÃO EM DEZ MINUTOS

No seu Diário de Fronteira, faça três colunas: "a IA faz bem", "a IA faz, mas eu preciso conferir" e "só a minha experiência resolve". Distribua nelas cinco tarefas do seu trabalho. Depois, escolha uma da primeira coluna para delegar esta semana e uma da terceira para fortalecer. Coloque a data.

Fim da Parte IV

Chegamos ao fim da quarta parte. Vimos empresas saírem do zero e chegarem a US$ 100 milhões em receita anualizada em meses, e aprendemos por que ARR não é lucro nem garantia. Conhecemos a empresa de uma pessoa, com o alcance de uma organização e a responsabilidade de um dono. Entendemos por que a inteligência fica mais barata para usar enquanto fica mais cara para construir. E, agora, vimos que competir com IA não é uma questão de valor humano, e sim de método, julgamento e oportunidade.

Entender a compressão dos negócios era o quarto passo. O quinto é olhar para as pessoas que vão viver dentro dela.

Porque, se os métodos de trabalho mudam, mudam também os caminhos para aprendê-los. Mudam a escola, a faculdade, o primeiro emprego, a contratação e a forma de provar o que se sabe fazer. O degrau de entrada que vimos se estreitar neste capítulo é só o começo dessa conversa.

Na Parte V, vamos aprender, trabalhar e contratar de novo. Começando por uma pergunta que muitas famílias já estão fazendo à mesa do jantar: a faculdade ainda importa?

Notas

1. Lütke, Tobi. Memorando interno da Shopify sobre uso de IA, divulgado publicamente em 7 abr. 2025. Ver: The Verge. "Shopify CEO says no new hires without proof AI can't do the job." Abr. 2025. https://www.theverge.com/news/644943/shopify-ceo-memo-ai-hires-job

2. Brynjolfsson, Erik; Li, Danielle; Raymond, Lindsey. "Generative AI at Work." The Quarterly Journal of Economics, v. 140, n. 2, maio 2025, p. 889–942 (publicado on-line em 4 fev. 2025). https://doi.org/10.1093/qje/qjae044; versão anterior: NBER Working Paper 31161 (2023), que reportava 14% em média e 34% para os menos experientes. https://www.nber.org/papers/w31161

3. Brynjolfsson, Erik; Chandar, Bharat; Chen, Ruyu. "Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence." Stanford Digital Economy Lab, versão revisada de ago. 2026 (dados ADP até jun. 2026). https://digitaleconomy.stanford.edu/news/canariesaug26/

4. Microsoft. "Global AI Diffusion – Q2 2026 Trends and Insights." 21 set. 2026. https://www.microsoft.com/en-us/research/wp-content/uploads/2026/09/Microsoft-AI-Diffusion-Report-2026-Q2.pdf; Microsoft On the Issues. "The continued state of global AI diffusion in 2026." 21 set. 2026. https://blogs.microsoft.com/on-the-issues/2026/09/21/the-continued-state-of-global-ai-diffusion-in-2026/

5. Cetic.br/NIC.br. "50 milhões de brasileiros já usam IA, mas potenciais benefícios continuam limitados às camadas de maior renda e escolaridade." 9 dez. 2025 (TIC Domicílios 2025). https://cetic.br/noticia/50-milhoes-de-brasileiros-ja-usam-ia-mas-potenciais-beneficios-continuam-limitados-as-camadas-de-maior-renda-e-escolaridade/

6. Dell'Acqua, Fabrizio et al. "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality." Harvard Business School Working Paper 24-013, set. 2023; versão publicada em Organization Science. https://www.hbs.edu/faculty/Pages/item.aspx?num=64700; https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4573321

7. Comissão Europeia. "Article 4: AI literacy." AI Act Service Desk. https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-4; Sapuppo, Valentina Grazia. "AI literacy: the Digital Omnibus rewrites Article 4 of the AI Act." Law & Technology, 30 jul. 2026. https://lawandtechnology.eu/en/ai-literacy-digital-omnibus-article-4-ai-act/; Regulamento (UE) 2026/1744 (Digital Omnibus on AI), Jornal Oficial, 24 jul. 2026.

PARTE V

Aprender, trabalhar e contratar de novo

CAPÍTULO 21

A faculdade ainda importa?

Tese do capítulo: a faculdade continua valiosa em média, mas o diploma deixou de ser uma aposta automaticamente boa; custo, campo, conclusão, experiência e capacidade de reaprender determinam seu retorno.

Terminei a Parte IV com uma pergunta que muitas famílias já estão fazendo à mesa do jantar. Ela merece começar a quinta parte deste livro, porque é nela que a revolução que vimos nos negócios encontra a vida de cada pessoa: a escolha de onde estudar, por quanto tempo e a que preço.

1. A pergunta que muita gente tem medo de fazer

Hoje, quando olho para uma faculdade tradicional, às vezes tenho a sensação de estar vendo uma estrutura enorme tentando preparar pessoas para um mundo que já começou a desaparecer. Quatro anos sempre foram muito tempo. Agora, em quatro anos, uma profissão pode mudar de ferramentas, linguagem, processo e até de lógica econômica várias vezes.

Foi por isso que eu fiz uma pergunta direta: faculdade hoje em dia ainda importa? E um PhD? O ensino precisa mudar porque, do jeito que está, parece uma perda de tempo absoluta para muita gente?

Eu gosto de perguntas diretas, mas não quero responder com uma frase irresponsável. Dizer que toda faculdade é perda de

tempo seria tão superficial quanto dizer que qualquer diploma continua sendo um passaporte seguro para o futuro. Os dados não sustentam nenhum desses extremos.

Nos Estados Unidos, pessoas com bacharelado ainda ganham mais e enfrentam menos desemprego, em média, do que pessoas com apenas ensino médio. Em 2025, a mediana semanal dos trabalhadores em tempo integral com 25 anos ou mais foi de US$ 1.578 para quem tinha somente bacharelado e US$ 966 para quem tinha apenas diploma do ensino médio.[1] Em setembro de 2026, o desemprego entre pessoas com bacharelado ou nível superior estava em 2,5%, contra 4,4% entre formados apenas no ensino médio.[2]

Portanto, a faculdade não deixou de ter valor. Mas a média nacional não paga a dívida de uma pessoa, não devolve quatro anos da vida de outra e não garante que um curso específico produza uma carreira. A média mistura universidades excelentes e fracas, cursos baratos e caríssimos, áreas de alta demanda e áreas com poucas vagas, estudantes que concluem e estudantes que ficam com dívida sem receber o diploma.

O mercado dos recém-formados mostra a tensão. No segundo trimestre de 2026, segundo o Federal Reserve Bank of New York, aproximadamente 5,6% dos formados de 22 a 27 anos estavam desempregados e 42% estavam subempregados, trabalhando em ocupações que normalmente não exigem formação universitária.[3] No Capítulo 20, vimos que é justamente no degrau de entrada que a IA começa a pesar mais.

O diploma ainda abre portas. A pergunta é se ele abre a porta que você precisa, pelo preço que você pode pagar, no tempo que você pode investir e para um trabalho que continuará existindo de forma reconhecível.

A pergunta certa não é "faculdade ou não faculdade?". É: para esta pessoa, para esta profissão, por este preço, diante de quais alternativas?

NÃO CONFUNDA

Média é o que acontece com o conjunto dos formados. *Promessa* é o que acontecerá com você. A primeira não garante a segunda.

Desemprego é não ter trabalho e estar procurando. *Subemprego*, no indicador do Fed de Nova York, é trabalhar numa ocupação que normalmente não exige diploma universitário.

Diploma exigido por lei é condição para exercer certas profissões, como medicina ou engenharia. *Diploma exigido por hábito* é um filtro que o empregador poderia trocar por uma prova séria de competência.

2. O diploma ainda paga, mas já não compra o que comprava

Durante grande parte do século passado, a universidade reuniu coisas que eram difíceis de obter em qualquer outro lugar: professores, bibliotecas, laboratórios, equipamentos, pesquisa, colegas ambiciosos, recrutadores, prestígio e uma sequência organizada de aprendizagem. Para muitas profissões, ela também entregava a credencial legal necessária para exercer o trabalho.

Algumas dessas funções continuam indispensáveis. Eu não quero um cirurgião que decidiu estudar medicina somente por vídeos. Não quero uma pessoa projetando uma ponte sem base matemática, laboratório, supervisão e responsabilidade profissional. Pesquisa científica original exige método, infraestrutura, crítica e tempo. Direito, enfermagem, engenharia licenciada e várias outras áreas possuem obrigações que não

desaparecem porque uma IA consegue escrever uma explicação bonita.

Mas a universidade perdeu o monopólio do acesso à informação e está perdendo o monopólio da instrução. Uma pessoa pode conversar com um tutor artificial, assistir a especialistas do mundo inteiro, praticar programação, criar um produto, aprender um idioma, analisar documentos e construir um portfólio sem esperar o início do próximo semestre.

Isso muda o cálculo. Antes, a alternativa à faculdade frequentemente era "não estudar". Agora, a alternativa pode ser uma combinação de curso técnico, certificado, aprendizagem prática, trabalho, comunidade, mentor, portfólio e tutor de IA. Essa combinação não serve para todas as pessoas nem para todas as profissões, mas já é real demais para ser tratada como exceção.

A pesquisa de retorno sobre investimento de 4.600 instituições feita por Georgetown mostra por que a resposta depende do horizonte. Programas mais curtos podem produzir retorno mais rápido porque custam menos e colocam o estudante no mercado antes. Muitos bacharelados recuperam vantagem ao longo de décadas. Porém, campo de estudo, preço líquido, conclusão e instituição mudam radicalmente o resultado.[4]

Também precisamos separar duas palavras que foram tratadas como sinônimas: educação e credencial. Educação é aquilo que você aprende, consegue aplicar e continua desenvolvendo. Credencial é o sinal formal de que passou por determinado caminho. Em alguns mercados, o sinal ainda é necessário. Em outros, ele funcionou durante anos como um atalho preguiçoso para recrutadores que não sabiam medir competência.

A inteligência artificial enfraquece esse atalho de duas formas. Primeiro, torna parte do conhecimento mais acessível fora da

instituição. Segundo, torna muito mais fácil produzir um currículo, uma carta, um ensaio e até um portfólio aparentemente brilhante. A consequência não deveria ser abandonar padrões. Deveria ser exigir evidência melhor.

A própria educação enfrenta o mesmo problema. Quando uma IA consegue redigir o trabalho entregue em casa, a nota deixa de provar que o estudante sabe pensar ou escrever. A OCDE alerta que o uso de IA pode melhorar a qualidade imediata de uma tarefa sem gerar aprendizagem duradoura. O estudante parece melhor enquanto usa a ferramenta, mas pode não manter o desempenho quando precisa trabalhar sem ela.[5,6]

Isso não é um argumento para proibir a IA. É um argumento para redesenhar a escola. Precisamos de mais defesa oral, laboratório, discussão, projetos, trabalho de campo, colaboração, prática, revisão e demonstração de processo. A pessoa deve poder usar IA, mas também precisa explicar o que aceitou, o que rejeitou, como verificou e o que aprendeu.

A IA não matou a faculdade. Ela matou a desculpa para uma faculdade cara, lenta e desconectada continuar ensinando como se nada tivesse mudado.

FOTOGRAFIA DA FRONTEIRA — EDUCAÇÃO E TRABALHO (6 DE OUTUBRO DE 2026)

Renda (BLS, 2025): mediana semanal de US$ 1.578 com bacharelado e US$ 966 só com ensino médio (tempo integral, 25 anos ou mais).

Desemprego (BLS, set. 2026): 2,5% com bacharelado ou mais; 4,4% só com ensino médio (com ajuste sazonal).

Recém-formados de 22 a 27 anos (Fed de Nova York, 2º tri. 2026): cerca de 5,6% de desemprego e 42% de subemprego.

Retorno (Georgetown, 2025): em 10 anos, certificados e cursos de dois anos costumam render mais; em 40 anos, o bacharelado quase sempre passa à frente. Nas instituições públicas, a mediana em 40 anos foi de cerca de US$ 1,8 milhão no bacharelado, contra US$ 1,43 milhão nos cursos de dois anos e US$ 1,37 milhão nos certificados.

Aprendizagem (OCDE, jan. 2026): num experimento na Turquia, o GPT-4 melhorou o desempenho imediato em matemática, mas, sem a ferramenta, os alunos que o usaram foram 17% pior do que os que nunca o tiveram.

Fatos vivos: confira o Atlas Vivo da IA.

3. A decisão que substitui o piloto automático

Durante décadas, muitas famílias seguiram uma sequência quase automática: terminar a escola, entrar em uma universidade, obter um diploma e procurar emprego. Essa sequência ainda pode ser correta. O problema é tratá-la como resposta antes de fazer a pergunta.

Uma decisão educacional na era da IA deveria começar por sete verificações.

1. A profissão exige legalmente um diploma, licença, laboratório ou treinamento supervisionado?
2. Qual é o custo total: mensalidade, juros, moradia, transporte e renda que a pessoa deixa de receber?
3. Qual é a chance real de concluir o programa, e o que acontece financeiramente se não concluir?
4. Quais são os resultados daquele curso e daquela instituição, não apenas a média de todos os formados?
5. O programa oferece prática, equipamentos, pacientes, pesquisa, mentores, estágio e rede que seriam difíceis de construir sozinho?
6. O estudante sairá com um portfólio, experiência e capacidade de trabalhar com IA, ou apenas com créditos acumulados?
7. Existe uma alternativa mais rápida e barata que produza o mesmo acesso, competência e credencial?

Essas perguntas não diminuem o valor do conhecimento. Elas aumentam a responsabilidade da escolha.

Para uma pessoa de 15 ou 18 anos, a resposta talvez seja universidade, mas com uma estratégia muito mais ativa: escolher custo e curso com cuidado, trabalhar cedo, construir portfólio, usar IA como tutora e não como substituta, procurar mentores e revisar o plano a cada semestre.

Para uma pessoa de 25 ou 30 anos, a resposta pode ser uma graduação, uma formação técnica, uma certificação, um projeto empresarial ou uma combinação modular. Não existe mérito em passar quatro anos obtendo uma credencial que o trabalho desejado não exige. Também não existe mérito em rejeitar uma formação necessária apenas porque alguém na internet ficou rico sem ela.

Para uma pessoa de 50 ou 60 anos, voltar a estudar pode ser extraordinário, mas não precisa começar com outra graduação. Pode começar com uma competência específica, um processo real e uma aplicação imediata. Experiência combinada com alfabetização em IA pode valer mais do que acumular conteúdo sem uso.

As universidades que prosperarem não serão aquelas que fingirem que ChatGPT, Claude ou Gemini não existem. Serão as que conseguirem combinar profundidade, comunidade, pesquisa, prática, ética e tecnologia; aquelas que provarem que seus estudantes aprendem a formular perguntas, construir, verificar, colaborar e assumir responsabilidade.

Os empregadores também precisam mudar. Exigir bacharelado para qualquer função administrativa não é rigor; muitas vezes é terceirizar a avaliação para uma instituição. Quando o diploma for indispensável, diga por quê. Quando não for, teste capacidade de maneira séria.

O futuro da educação não será simplesmente universidade contra internet, professor contra IA ou diploma contra portfólio. Será uma combinação em que cada componente terá de justificar o valor que entrega.

A educação continuará sendo uma das melhores formas de ampliar uma vida. Mas o formato da educação não recebe o direito de permanecer igual apenas porque um dia funcionou.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: a faculdade pode ser um ótimo caminho, mas não é um trilho automático. Antes de escolher, descubra o que a profissão que você quer realmente exige, quanto o curso custa e o que você vai conseguir mostrar quando sair, além do diploma.

Para quem tem 65 anos: na sua época, o diploma muitas vezes bastava para abrir a porta e garantir a carreira. Hoje ele ainda ajuda, mas os netos precisam também de prática, portfólio e capacidade de reaprender. Ajudá-los a fazer as perguntas certas vale mais do que repetir que "faculdade é garantia".

O QUE AINDA NÃO SABEMOS

Se o aumento do subemprego e do desemprego entre recém-formados é temporário ou um sinal duradouro da IA no degrau de entrada. Quanto tempo as universidades vão levar para redesenhar avaliação e currículo. Se os empregadores vão realmente trocar o diploma por provas de competência. E se as alternativas mais curtas vão manter o retorno quando muito mais gente seguir por elas.

O que mudou / O que significa / O que fazer agora

O que mudou: informação e prática inicial ficaram mais acessíveis, enquanto o mercado de entrada ficou mais competitivo e a qualidade dos diplomas passou a variar mais.

O que significa: o valor passa a depender de aprendizagem real, credencial necessária, experiência, rede, custo e capacidade de atualização.

O que fazer agora: antes de assumir dívida ou investir anos, compare profissão, custo, resultados, alternativas e competências. Trate a escolha como investimento, não como ritual social. A IA não tornou a faculdade inútil; tornou uma faculdade cara, genérica e desconectada do trabalho muito mais difícil de justificar.

AÇÃO EM DEZ MINUTOS

No seu Diário de Fronteira, escolha um curso que você, um filho, um neto ou alguém da sua equipe esteja considerando. Responda às sete verificações deste capítulo, uma linha para cada. Para os números, procure os resultados daquele curso e daquela instituição: nos Estados Unidos, no College Scorecard; no Brasil, no cadastro e-MEC e no Conceito Enade, do Inep.[7] Coloque a data e marque as perguntas que ficaram sem resposta.

Mas a pergunta que fiz no começo deste capítulo tinha uma segunda parte: e um PhD? Se a faculdade deixou de ser uma aposta automática, o que acontece com o nível mais alto da formação acadêmica, quando a própria inteligência começa a ficar abundante? É o que vamos ver no próximo capítulo.

Notas

1. U.S. Bureau of Labor Statistics. "Median weekly earnings of full-time wage and salary workers 25 years and over by sex and educational attainment" (Current Population Survey, médias anuais de 2025). https://www.bls.gov/cps/cpsaat37b.htm; BLS. "Usual Weekly Earnings of Wage and Salary Workers – Fourth Quarter 2025", tabela 9. 28 jan. 2026. https://www.bls.gov/news.release/archives/wkyeng_01282026.htm

2. U.S. Bureau of Labor Statistics. "The Employment Situation – September 2026", tabela A-4, "Employment status of the civilian population 25 years and over by educational attainment" (dados com ajuste sazonal). 2 out. 2026. https://www.bls.gov/news.release/empsit.t04.htm

3. Federal Reserve Bank of New York. "The Labor Market for Recent College Graduates." Destaques de 2026:Q2 (recém-formados de 22 a 27 anos com bacharelado ou mais). https://www.newyorkfed.org/research/college-labor-market

4. Cheah, Ban; Van Der Werf, Martin; Morris, Catherine; Strohl, Jeff. "Ranking 4,600 Colleges by ROI (2025)." Georgetown University Center on Education and the Workforce, 26 fev. 2025 (dados do College Scorecard). https://cew.georgetown.edu/cew-reports/roi2025/; CEW. "CEW's Year in Review: 2025." https://cew.georgetown.edu/resource/cew-2025-blog-year-in-review/

5. OECD. "OECD Digital Education Outlook 2026: Exploring Effective Uses of Generative AI in Education." 19 jan. 2026. https://www.oecd.org/en/publications/oecd-digital-education-outlook-2026_062a7394-en.html

6. OECD. "What should teachers teach and students learn in a future of powerful AI?" OECD Education Spotlights n. 20, 23 maio 2025. https://www.oecd.org/en/publications/what-should-teachers-teach-and-students-learn-in-a-future-of-powerful-ai_ca56c7d6-en.html

7. U.S. Department of Education. College Scorecard. https://collegescorecard.ed.gov/; Ministério da Educação. Cadastro Nacional de Cursos e Instituições de Educação Superior (e-MEC). https://emec.mec.gov.br/; Inep. Exame Nacional de Desempenho dos Estudantes (Enade). https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enade

CAPÍTULO 22

PhD na era da inteligência abundante

Tese do capítulo: um doutorado não é obsoleto quando o mundo precisa de pesquisa original; o que perdeu valor é usar o título como substituto de curiosidade, resultado ou capacidade de trabalhar com ferramentas novas.

Terminei o capítulo anterior lembrando que a minha pergunta tinha uma segunda parte: e um PhD?

Ela nasceu de uma comparação simples. Nos Estados Unidos, quem concluiu um doutorado de pesquisa em 2024 levou, em mediana, 7,3 anos desde a entrada na pós-graduação, e 5,7 anos só dentro do programa de doutorado.[1] Ao mesmo tempo, em fevereiro de 2025, a OpenAI lançou o Deep Research, um agente que, segundo a empresa, faz em dezenas de minutos uma pesquisa que levaria muitas horas para uma pessoa, lendo e cruzando centenas de fontes.[2] E, no Capítulo 19, vimos que acertar 75% de uma prova de ciências em nível de doutorado passou a custar quatro centésimos de centavo de dólar por questão.

Então a pergunta ficou inevitável: se uma ferramenta percorre milhares de documentos em minutos, sete anos de formação continuam justificáveis?

A minha resposta curta é: depende do que o doutorado está treinando. Se ele treina acesso a informação, a IA já comprimiu boa parte dessa vantagem. Se ele treina a capacidade de fazer

perguntas novas e provar respostas difíceis, o mundo vai precisar mais dele, não menos.

1. O que um PhD deveria treinar: formular perguntas, método, evidência, profundidade e resistência à incerteza

Um doutorado nunca deveria ter sido apenas "saber muito sobre um assunto". O que ele deveria ensinar é outra coisa: escolher uma pergunta que ninguém respondeu, desenhar um método para testá-la, distinguir evidência de impressão, ir fundo o bastante para perceber o que os outros deixaram passar e aguentar anos de incerteza sem desistir nem se enganar.

Essas capacidades nunca foram tão necessárias, e um exemplo brasileiro mostra por quê. A Iniciativa Brasileira de Reprodutibilidade, financiada pelo Instituto Serrapilheira, reuniu 56 laboratórios do país para repetir experimentos biomédicos publicados por grupos brasileiros, usando três técnicas comuns de laboratório. O projeto levou cerca de sete anos. Na versão mais recente do estudo, revisada pela plataforma eLife em julho de 2026, foram consideradas válidas 90 repetições de 45 experimentos. Dependendo do critério usado, só entre 20% e 44% dos resultados se confirmaram. Em mediana, os efeitos encontrados nas repetições foram 58% menores do que nos artigos originais.[3]

Isso não quer dizer que a ciência brasileira seja pior do que a de outros países. Os revisores observaram que as taxas são parecidas com as de outros grandes esforços de replicação na biomedicina.[3] Quer dizer que publicar não é o mesmo que provar. E essa diferença, entre um resultado que parece verdadeiro e um

que resiste a ser testado de novo, é exatamente o que um bom doutorado deveria ensinar a enxergar.

A IA, aliás, pode ampliar enormemente quem tem esse treinamento. No Capítulo 9, vimos o AlphaFold, que rendeu a Demis Hassabis e John Jumper metade do Nobel de Química de 2024. Uma ferramenta extraordinária, criada e validada por pessoas com anos de profundidade científica. A IA não dispensou o método. Ela deu ao método um alcance novo.

NÃO CONFUNDA

Informação é saber o que já foi publicado. A IA tornou esse acesso rápido e barato.

Pesquisa original é produzir conhecimento que ainda não existe e provar que ele se sustenta. Isso continua exigindo pergunta, método, evidência e tempo.

Título é o sinal de que alguém passou por esse treinamento. *Capacidade* é continuar fazendo isso depois dele.

2. O que a IA comprime: revisão de literatura, rascunho, código, análise e acesso ao conhecimento

Agora o outro lado. Boa parte do trabalho que ocupava os anos de um doutorando está sendo comprimida.

A revisão de literatura, que podia consumir meses, ganhou agentes como o Deep Research. A escrita também mudou: um estudo publicado em julho de 2025 na revista Science Advances analisou mais de 15 milhões de resumos de artigos biomédicos e estimou que pelo menos 13,5% dos publicados em 2024 tinham

passado por um modelo de linguagem.[4] Código e análise de dados, que exigiam semanas de tentativa e erro, hoje saem em rascunho em minutos.

E o movimento chegou à pesquisa inteira. Em março de 2026, a revista Nature publicou o artigo que descreve o AI Scientist, sistema da startup japonesa Sakana AI desenvolvido com pesquisadores da University of British Columbia, do Vector Institute e da Universidade de Oxford. Segundo a empresa, o sistema gera ideias, busca e lê a literatura, programa e executa experimentos e escreve o artigo inteiro. Um dos artigos produzidos assim passou pela revisão às cegas de um workshop da conferência ICLR 2025, com nota média de 6,33 em 10, e foi retirado antes da publicação, como combinado com os organizadores. A própria Sakana reconhece que o sistema ainda produz ideias ingênuas, tem dificuldade com rigor metodológico e pode inventar citações.[5] No Capítulo 14, vimos que Jeff Dean e outros veteranos deixaram o Google para fundar a Discovery Loop, com a meta declarada de automatizar a pesquisa em aprendizado de máquina, ciência e engenharia. Ou seja: a automação da pesquisa deixou de ser curiosidade de laboratório e virou plano de negócios.

Mas compressão sem julgamento cobra seu preço. Em janeiro de 2026, a startup GPTZero analisou os 4.841 artigos aceitos pela NeurIPS 2025, uma das conferências mais prestigiadas de IA do mundo, e encontrou 100 citações inventadas em 51 deles. Todos tinham passado por revisão de especialistas.[6] É pouco no total, mas o recado é forte: se os maiores especialistas em IA deixam escapar referências falsas, ninguém está imune.

As instituições começaram a reagir. Em julho de 2025, os Institutos Nacionais de Saúde dos Estados Unidos, o NIH, principal financiador público de pesquisa biomédica do país, anunciaram que não considerariam pedidos de financiamento

desenvolvidos substancialmente por IA, por não representarem ideias originais dos pesquisadores. O órgão também passou a limitar cada pesquisador a seis pedidos por ano, depois de identificar casos de mais de 40 pedidos distintos numa única rodada, alguns possivelmente gerados com IA.[7]

A IA pode escrever a tese. Não pode responder por ela.

3. Quem deve fazer um doutorado agora, e quem não deveria usá-lo como abrigo contra um mercado em transformação

O doutorado ainda abre portas, mas não da mesma forma em todas as áreas. Nos Estados Unidos, segundo a pesquisa anual com todos os novos doutores, 70% dos que concluíram doutorado em ciências e engenharias em 2024 e responderam à pergunta já tinham emprego ou pós-doutorado garantido ao se formar. A taxa ia de 63% nas ciências biológicas e biomédicas a 80% em psicologia, e caiu em todas as grandes áreas de 2023 para 2024.[8] Entre os que tinham emprego definido no país, fora o pós-doutorado, 40% iam para a academia e 40% para a indústria.[8]

Os salários contam outra parte da história. Em todas as áreas, o salário mediano esperado na indústria era maior do que na academia ou no pós-doutorado. A mediana na indústria ia de US$ 180 mil por ano em ciência da computação a US$ 97 mil em psicologia. No pós-doutorado, ficava entre US$ 58 mil e US$ 70 mil.[9]

E no Brasil? O estudo Mestres e Doutores 2024, do CGEE, mostra que o país titulou 319.211 doutores entre 1996 e 2021. Em 2021, havia 4,4 doutores para cada mil empregos formais, contra

12,9 na média da União Europeia em 2022.[10] Em outras palavras, o Brasil não tem doutores demais. O desafio é que esses doutores encontrem problemas à altura do que aprenderam, dentro e fora da universidade.

Repare que nenhum desses números diz que o doutorado deixou de valer. Eles dizem que o valor depende muito da área, do destino e do que a pessoa faz com o título. Diante disso, minha resposta à pergunta do início é esta.

Um doutorado faz sentido para quem tem um problema que exige profundidade: uma pergunta científica que ninguém respondeu, uma carreira que exige o título, como a pesquisa acadêmica, ou uma área em que a indústria contrata justamente quem sabe produzir conhecimento novo. Faz ainda mais sentido para quem chega disposto a usar a IA como laboratório ampliado, e não como atalho.

Um doutorado não faz sentido como abrigo. Se a principal razão é adiar a entrada num mercado que assusta, ganhar tempo para decidir o que fazer ou acumular um título porque ele parece garantir respeito, sete anos podem se transformar no caminho mais caro para chegar atrasado a um mundo que mudou.

FOTOGRAFIA DA FRONTEIRA — DOUTORADO E IA (6 DE OUTUBRO DE 2026)

Estados Unidos (dados de 2024, publicados em 2026): 58.131 doutorados de pesquisa; mediana de 7,3 anos desde a entrada na pós-graduação; 70% dos novos doutores em ciências e engenharias com emprego ou pós-doutorado garantido.

Salários esperados (2024): indústria de US$ 97 mil a US$ 180 mil por ano, conforme a área; pós-doutorado de US$ 58 mil a US$ 70 mil.

Brasil (CGEE, dados até 2021): 319.211 doutores titulados desde 1996; 4,4 doutores por mil empregos formais.

Pesquisa automatizada: AI Scientist, da Sakana AI, publicado na Nature em mar. 2026.

Integridade: 100 citações inventadas em 51 artigos aceitos pela NeurIPS 2025; NIH recusa pedidos desenvolvidos substancialmente por IA desde set. 2025.

Fatos vivos: confira o Atlas Vivo da IA.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: fazer doutorado é como virar detetive de uma pergunta que ninguém conseguiu resolver. A IA pode ser a sua lupa mais poderosa, mas não escolhe o caso por você, nem prova quem tem razão.

Para quem tem 65 anos: antes, o doutor era quem tinha lido tudo sobre um assunto. Hoje, qualquer pessoa pode pedir a uma máquina que leia tudo em minutos. O valor do doutor passou a estar em saber o que perguntar, como testar e quando desconfiar.

O QUE AINDA NÃO SABEMOS

Se sistemas como o AI Scientist vão produzir descobertas realmente novas fora da computação, em laboratórios com experimentos físicos. Se os doutorados vão ficar mais curtos à medida que a IA comprime revisão, escrita e análise. Como universidades e revistas vão separar contribuição humana de contribuição da máquina. E se a revisão por pares vai aguentar o volume de artigos que a IA torna possível produzir.

O que mudou / O que significa / O que fazer agora

O que mudou: a IA comprimiu partes do trabalho de pesquisa que consumiam anos, como revisão de literatura, rascunho, código e análise, e já produz artigos capazes de passar por revisão em alguns contextos.

O que significa: o valor do doutorado se desloca do acúmulo de informação para a capacidade de formular perguntas, testar com rigor e responder pelo resultado. Ao mesmo tempo, integridade e reprodutibilidade ficam mais importantes, porque ficou mais fácil produzir algo que parece ciência.

O que fazer agora: antes de começar um doutorado, escreva qual problema você quer resolver e por que ele exige anos de profundidade. Verifique para onde vão e quanto ganham os doutores da sua área. E, se decidir seguir, aprenda a usar a IA como laboratório ampliado, conferindo cada fonte e cada número. O PhD vale quando o problema exige profundidade; não quando serve apenas para adiar a entrada no mundo que mudou.

AÇÃO EM DEZ MINUTOS

No seu Diário de Fronteira, peça a uma ferramenta de IA um resumo do que se sabe sobre um tema do seu trabalho, com cinco referências. Depois, abra cada uma delas e confira: o artigo existe? Diz o que a IA afirmou? Anote quantas passaram no teste. Coloque a data.

Mas o desafio da formação não começa na universidade. Se a faculdade e o doutorado precisam se reinventar, a escola, onde tudo começa, enfrenta um problema ainda maior. No próximo capítulo, vamos olhar para a escola que continua ensinando para um mundo que terminou.

Notas

1. National Center for Science and Engineering Statistics (NCSES). "Doctorate Recipients from U.S. Universities: 2024." NSF 26-315, 13 mar. 2026. https://ncses.nsf.gov/pubs/nsf26315; tabela 1-12, "Median years to research doctorate, by trend broad field of doctorate: Selected years, 1974–2024." https://www.ncses.nsf.gov/pubs/nsf25349/assets/data-tables/tables/nsf25349-tab001-012.pdf

2. OpenAI. "Introducing deep research." 2 fev. 2025. https://openai.com/index/introducing-deep-research/

3. Brazilian Reproducibility Initiative. "Estimating the replicability of Brazilian biomedical science." bioRxiv, preprint (v5), doi:10.1101/2025.04.02.645026. https://www.biorxiv.org/content/10.1101/2025.04.02.645026v5; eLife, reviewed preprint 111001, 10 jul. 2026. https://elifesciences.org/reviewed-preprints/111001v1/reviews; eLife. "Reproducibility: How replicable is biomedical science in Brazil?" (entrevista). https://elifesciences.org/interviews/388d5959/brazilian-reproducibility-initiative

4. Kobak, Dmitry et al. "Delving into LLM-assisted writing in biomedical publications through excess vocabulary." Science Advances, 2 jul. 2025. https://doi.org/10.1126/sciadv.adt3813; preprint: https://arxiv.org/abs/2406.07016

5. Sakana AI. "The AI Scientist: Towards Fully Automated AI Research, Now Published in Nature." 26 mar. 2026. https://sakana.ai/ai-scientist-nature/; Nature. "Towards end-to-end automation of AI research." 2026. https://www.nature.com/articles/s41586-026-10265-5

6. Bort, Julie. "Irony alert: Hallucinated citations found in papers from NeurIPS, the prestigious AI conference." TechCrunch, 21 jan. 2026. https://techcrunch.com/2026/01/21/irony-alert-hallucinated-citations-found-in-papers-from-neurips-the-prestigious-ai-conference/; GPTZero. "GPTZero finds 100 new hallucinations in NeurIPS 2025 accepted papers." https://gptzero.me/news/neurips/

7. National Institutes of Health. "NOT-OD-25-132: Supporting Fairness and Originality in NIH Research Applications." Jul. 2025 (válido a partir de 25 set. 2025). https://grants.nih.gov/grants/guide/notice-files/NOT-OD-25-132.html; NIH Extramural Nexus. "Apply Responsibly: Policy on AI Use in NIH Research Applications and Limiting Submissions per PI." Jul. 2025. https://www.grants.nih.gov/news-events/nih-extramural-nexus-news/2025/07/apply-responsibly-policy-on-ai-use-in-nih-research-applications-and-limiting-submissions-per-pi

8. NCSES. "Doctorate Recipients from U.S. Universities: 2024", seção "Postgraduation Trends." NSF 26-315, 13 mar. 2026. https://ncses.nsf.gov/pubs/nsf26315/report/postgraduation-trends

9. NCSES. "Expected Median Salaries for Doctorate Recipients Are Highest in the Industry Sector." NSF 26-312, 17 fev. 2026. https://ncses.nsf.gov/pubs/nsf26312

10. Centro de Gestão e Estudos Estratégicos (CGEE). "Brasil: Mestres e Doutores 2024" (apresentação e seção 3.1, "Número de empregados"; dados de 1996 a 2021). https://mestresdoutores2024.cgee.org.br/-/apresentacao; https://mestresdoutores2024.cgee.org.br/-/3.1-numero-empregados

CAPÍTULO 23

A escola que ensina para um mundo que terminou

Tese do capítulo: quando a IA consegue produzir a lição de casa, a escola precisa voltar a provar que está produzindo aprendizado.

Terminei o capítulo anterior dizendo que o desafio da formação não começa na universidade. Começa muito antes, numa sala de aula com carteiras enfileiradas, um quadro na frente e uma lição de casa para entregar na semana seguinte.

Quando penso nessa sala, penso numa expressão em inglês: *legacy structures*. São as estruturas herdadas, criadas para um mundo que já existiu e que continuam de pé porque sempre estiveram ali. Uma grande parte da sociedade está ancorada nelas. Empresas, governos, carreiras e, talvez mais do que qualquer outra instituição, a escola.

E aqui está a pergunta que mais me incomoda: como sair dessas estruturas sem destruir o que ainda funciona?

Os números brasileiros mostram que a mudança já entrou na escola, com ou sem permissão. Segundo a pesquisa TIC Educação, do Cetic.br, feita entre agosto de 2024 e março de 2025, sete em cada dez alunos do ensino médio que usam a internet já recorriam à IA generativa para pesquisas escolares. Só 32% disseram ter recebido alguma orientação da escola sobre como usá-la.[1] No Capítulo 1, vimos que, nos Estados Unidos, mais da metade dos

adolescentes já usava chatbots para trabalhos escolares. O aluno chegou antes da escola.

1. Memorização, ensaio feito em casa e currículo estático diante de sistemas que respondem e escrevem

Durante gerações, a escola funcionou com três peças que se apoiavam umas nas outras.

A primeira era a memorização. Saber de cor datas, fórmulas e capitais fazia sentido quando a informação era escassa e cara, guardada em enciclopédias, como lembrei no Capítulo 3.

A segunda era o trabalho feito em casa, como a redação, o resumo ou a pesquisa. Ele servia como prova de que o aluno tinha lido, entendido e organizado o pensamento. O professor via o produto e deduzia o processo.

A terceira era o currículo estático, revisado de tempos em tempos, pensado para durar anos.

A IA quebrou as três ao mesmo tempo. A memória virou consulta instantânea. O trabalho de casa pode ser escrito por uma máquina em segundos, com uma qualidade que muitas vezes supera a do aluno. E o currículo envelhece mais rápido do que qualquer comissão consegue revisá-lo.

O problema mais sério está na segunda peça. Quando o produto pode ser fabricado sem o processo, ele deixa de provar o processo. O Digital Education Outlook 2026, da OCDE, publicado em janeiro de 2026, resume o risco: sem orientação pedagógica, terceirizar tarefas para a IA melhora o desempenho sem gerar ganho real de aprendizagem. A OCDE vai além. Quando a IA elimina o "esforço produtivo", aquela dificuldade que faz o

aprendizado se fixar, o aluno termina mais rápido e com resultado melhor, mas entende menos. Pode perder fôlego cognitivo, leitura profunda, atenção e perseverança. Os pesquisadores dão um nome a isso: "preguiça metacognitiva".[2] No Capítulo 21, vimos um exemplo concreto desse efeito num experimento com matemática na Turquia.

Em outras palavras, a nota continua chegando, mas pode não estar medindo mais nada.

NÃO CONFUNDA

Desempenho é a qualidade do que o aluno entrega hoje, com a ferramenta ao lado.

Aprendizagem é o que fica com ele amanhã, quando a ferramenta não estiver lá.

A IA pode melhorar o primeiro e prejudicar o segundo. Uma escola que mede só o desempenho pode estar aplaudindo exatamente o que deveria preocupá-la.

2. O que não pode ser terceirizado: base factual, raciocínio, escrita, debate, relações, ética, corpo e prática

Se a máquina responde e escreve, o que a escola ainda precisa ensinar? A minha lista é a do título, e cada item tem uma razão.

Base factual. Parece contraditório, mas, quanto mais a IA responde, mais conhecimento precisamos ter. No Capítulo 3, vimos que a IA pode inventar com total segurança. Só percebe o erro quem sabe o suficiente para desconfiar.

Raciocínio e escrita. Escrever não é só registrar o que se pensa. É uma forma de pensar. Em junho de 2025, pesquisadores do MIT Media Lab publicaram um estudo, ainda sem revisão por pares, com 54 participantes que escreveram redações em três grupos: com ChatGPT, com buscador ou sem nenhuma ferramenta. Os sinais de atividade cerebral, medidos por eletroencefalograma, foram mais fortes e mais distribuídos no grupo sem ferramentas e mais fracos no grupo que usou o ChatGPT. Quem usou o modelo se sentiu menos dono do próprio texto e teve dificuldade de citar corretamente o que tinha acabado de escrever.[3] É um estudo pequeno, e os próprios autores pedem mais pesquisa. Mas a intuição que ele reforça é antiga: quem não escreve, não fixa.

Debate e relações. Aprender a discordar, ouvir, defender uma ideia diante de outras pessoas e mudar de opinião com dignidade. Nenhuma conversa com uma máquina substitui uma sala cheia de gente diferente.

Ética. Saber o que é honesto, o que é autoria, o que é justo. Isso se aprende na prática, com exemplos e consequências.

Corpo e prática. Laboratório, oficina, esporte, arte, trabalho com as mãos, presença física. O Brasil fez aqui uma escolha clara. Em janeiro de 2025, a Lei 15.100 restringiu o uso de celulares pelos alunos na educação básica, inclusive no recreio, com exceções para uso pedagógico, acessibilidade e saúde.[4] No balanço divulgado pelo MEC em 30 de junho de 2026, 92% das escolas já tinham implementado a lei. Segundo os gestores, a medida ampliou a participação dos alunos nas atividades (97%) e a socialização presencial e a concentração (95%). E mais da metade das escolas públicas ampliou as ações de educação digital e midiática em 2025.[5] São percepções de quem dirige as escolas, não medidas de aprendizagem. Mas mostram que tirar a tela da mão não é o mesmo que tirar a tecnologia da escola.

A escola não precisa competir com a máquina em dar respostas. Precisa formar quem sabe o que fazer com elas.

3. Uma nova arquitetura: tutor de IA, professor, projetos, defesa oral, portfólio e aprendizagem ao longo da vida

Agora a parte que me enche de esperança. A mesma IA que ameaça a lição de casa pode se tornar o melhor tutor que muitos alunos já tiveram, desde que seja desenhada para ensinar, e não para responder.

As evidências estão aparecendo. Na Universidade Harvard, um experimento controlado com 194 alunos de física, publicado em junho de 2025 na revista Scientific Reports, comparou um tutor de IA, criado com base em boas práticas de ensino, com aulas presenciais de aprendizagem ativa. Os ganhos de aprendizagem de quem estudou com o tutor foram mais que o dobro, em menos tempo.[6] Na Nigéria, um programa piloto avaliado pelo Banco Mundial usou o Microsoft Copilot em seis semanas de aulas de inglês depois do horário escolar, sempre com professores guiando. Os alunos melhoraram o equivalente a 0,31 desvio-padrão, e o próprio Banco Mundial estimou o ganho como algo entre um ano e meio e dois anos de escola comum.[7]

Repare no detalhe que une os dois casos: a IA foi desenhada com intenção pedagógica, e havia um professor no centro. A OCDE chega à mesma conclusão. Tutores com pouca experiência que usaram apoio de IA elevaram em 9 pontos percentuais a taxa de aprovação dos alunos, e professores de ciências na Inglaterra reduziram em 31% o tempo de planejamento. Em 2024, 37% dos

professores dos países pesquisados pela organização já usavam IA generativa no trabalho.[2]

A Estônia transformou isso em política nacional. O programa AI Leap chegou, no ano letivo 2025–2026, a 154 das 156 escolas de ensino médio do país, com cerca de 5.000 professores e 20.000 alunos. Os alunos ganharam um aplicativo-tutor, o ITI, desenvolvido com pesquisadores estonianos e a OpenAI, que faz perguntas em vez de entregar a resposta pronta. Alguns alunos reclamaram que ele "irrita" justamente por isso. Para o diretor do programa, a mudança mais importante foi outra: as escolas começaram a discutir para que servem a lição de casa, o feedback e a avaliação.[8]

Essa é a nova arquitetura que eu imagino. O tutor de IA cuida da prática individual, no ritmo de cada aluno. O professor deixa de ser só transmissor e passa a ser arquiteto da aprendizagem, mentor e juiz do que é verdadeiro. Os projetos colocam o conhecimento em uso. A defesa oral, simples e poderosa, mostra quem entendeu: explique o que fez, o que aceitou da IA, o que rejeitou e como verificou. O portfólio registra o caminho, não só o resultado. E a aprendizagem ao longo da vida deixa de ser slogan, porque ninguém mais termina de aprender aos 18 ou aos 22 anos.

As instituições começaram a desenhar as regras. A UNESCO publicou, em setembro de 2024, referenciais de competências em IA para alunos e para professores.[9] Em junho de 2026, a OCDE e a Comissão Europeia lançaram a versão final de um referencial de alfabetização em IA para o ensino fundamental e médio, organizado em quatro domínios: interagir com a IA, criar com ela, gerenciá-la e ajudar a moldá-la. Ele alimenta o PISA de 2029, que vai avaliar a alfabetização midiática e em IA dos estudantes.[10]

E o Brasil? Em 1º de setembro de 2026, o Conselho Nacional de Educação aprovou as primeiras diretrizes nacionais para o uso de IA na educação. Até o 5º ano do ensino fundamental, as crianças não poderão usar ferramentas de IA de forma autônoma, apenas em atividades planejadas e mediadas por professores. O texto classifica os usos em quatro níveis de risco, do baixo ao proibido, e trata a IA como objeto de aprendizagem em toda a educação básica. O parecer ainda precisa ser homologado pelo ministro da Educação, e até a data desta fotografia eu não encontrei registro de que isso tenha acontecido. Depois da homologação, as instituições terão 12 meses para se adaptar.[11]

Se eu tivesse uma IA como professora particular aos 15 anos, lá em Mantena, eu sei exatamente o que teria pedido para ela me ensinar: como ser uma CEO numa empresa multinacional e como influenciar algo para a melhoria da humanidade. Era um sonho grande para uma menina do interior de Minas Gerais, filha de uma professora de primeiro grau e de um viajante que vendia camisas pelo interior da Bahia. Ainda adolescente, eu já administrava a minha própria conta bancária. Fui a primeira da minha família a terminar a faculdade. Anos depois, eu me tornei CEO.

FOTOGRAFIA DA FRONTEIRA — ESCOLA E IA (6 DE OUTUBRO DE 2026)

Brasil: 70% dos alunos do ensino médio usavam IA generativa em pesquisas escolares; só 32% receberam orientação (TIC Educação 2024). Lei 15.100 implementada em 92% das escolas (MEC, jun. 2026). Diretrizes de IA aprovadas pelo CNE em 1º set. 2026, aguardando homologação.

Tutores de IA: ganhos de aprendizagem mais que dobrados em física em Harvard (2025); +0,31 desvio-padrão em seis semanas na Nigéria (Banco Mundial, 2025).

Estônia: AI Leap em 154 de 156 escolas de ensino médio no primeiro ano, ampliado a todo o ensino médio em 2026–2027.

Referenciais: UNESCO (2024); OCDE e Comissão Europeia (jun. 2026), base do PISA 2029.

Fatos vivos: confira o Atlas Vivo da IA.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: usar a IA para fazer a lição é como pedir a alguém para ir à academia no seu lugar. O exercício fica feito, mas o músculo é de outra pessoa. Use a IA como treinador, não como substituto.

Para quem tem 65 anos: a escola em que você estudou ensinava a guardar informação, porque ela era difícil de encontrar. A escola de hoje precisa ensinar a julgar informação, porque ela está em toda parte, inclusive quando está errada.

O QUE AINDA NÃO SABEMOS

Se os bons resultados dos tutores de IA se mantêm em escala, por anos, e em escolas com pouca estrutura. Que efeito o uso intenso de IA na infância e na adolescência terá sobre memória, atenção e escrita. Como as escolas brasileiras vão formar professores a tempo de aplicar as novas diretrizes. E como provas e vestibulares vão medir aprendizagem quando a IA estiver em todo lugar.

O que mudou / O que significa / O que fazer agora

O que mudou: a IA passou a produzir, em segundos, a lição de casa, a redação e a pesquisa que a escola usava como prova de aprendizado. Ao mesmo tempo, tutores de IA bem desenhados começaram a mostrar ganhos reais de aprendizagem.

O que significa: a escola não perdeu a razão de existir, mas perdeu a desculpa para avaliar só o produto. Precisa voltar a observar o processo: o raciocínio, a conversa, a prática, a explicação de quem aprendeu.

O que fazer agora: se você é pai, mãe, avô, avó ou professor, pergunte menos "você fez a lição?" e mais "me explica como você chegou nisso". Peça para ver o caminho, inclusive o que a IA sugeriu e o que foi rejeitado. Se você lidera uma escola ou uma empresa que forma pessoas, troque parte das tarefas para casa por projetos, defesas orais e portfólios. A IA não torna a educação dispensável. Torna a educação sem aprendizagem indefensável.

AÇÃO EM DEZ MINUTOS

Escolha um assunto que você estudou na escola e peça a uma IA que seja seu tutor, com uma regra: ela não pode dar a resposta, só fazer perguntas e dar pistas. Depois de dez minutos, anote no seu Diário de Fronteira o que você lembrou, o que aprendeu e se preferiu esse jeito de estudar. Coloque a data.

E se a escola não consegue mais confiar na lição de casa para saber quem aprendeu, as empresas enfrentam o mesmo problema com currículos, cartas de apresentação e testes feitos em casa. No próximo capítulo, vamos ver como contratar quando todos podem parecer excelentes no papel.

Notas

1. Cetic.br/NIC.br. "Sete em cada dez alunos do Ensino Médio usam IA generativa em pesquisas escolares, revela TIC Educação." 16 set. 2025 (TIC Educação 2024; coleta de ago. 2024 a mar. 2025, com 7.476 alunos de 1.023 escolas). https://cetic.br/pt/noticia/sete-em-cada-dez-alunos-do-ensino-medio-usam-ia-generativa-em-pesquisas-escolares-revela-tic-educacao/

2. OCDE. "OECD Digital Education Outlook 2026: Exploring Effective Uses of Generative AI in Education." Jan. 2026. https://www.oecd.org/en/publications/oecd-digital-education-outlook-2026_062a7394-en.html; OCDE. "How to effectively use Generative AI in education." Jan. 2026. https://www.oecd.org/en/blogs/2026/01/how-to-effectively-use-generative-ai-in-education.html

3. Kosmyna, Nataliya et al. "Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task." MIT Media Lab, preprint arXiv:2506.08872, 10 jun. 2025. https://www.media.mit.edu/publications/your-brain-on-chatgpt/

4. Brasil. "Lei nº 15.100, de 13 de janeiro de 2025." Diário Oficial da União. https://www.in.gov.br/en/web/dou/-/lei-n-15.100-de-13-de-janeiro-de-2025-606772935

5. Ministério da Educação. "1º ano da Lei nº 15.100/2025" (relatório executivo; pesquisa com gestores de 8.189 escolas, mar.–abr. 2026). https://www.gov.br/mec/pt-br/celular-escola/arquivos/relatorio-executivo.pdf; Poder360. "Lei que controla celulares em escolas tem adesão de 92%, diz MEC." 30 jun. 2026. https://www.poder360.com.br/poder-educacao/lei-que-controla-celulares-em-escolas-tem-adesao-de-92-diz-mec/

6. Kestin, Gregory et al. "AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting." Scientific Reports 15, 17458, jun. 2025. https://doi.org/10.1038/s41598-025-97652-6

7. De Simone, Martín et al. "From Chalkboards to Chatbots: Evaluating the Impact of Generative AI on Learning Outcomes in Nigeria." World Bank Policy Research Working Paper 11125, maio 2025. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099548105192529324

8. AI Leap (TI-Hüpe). "AI Leap's First Year Led Schools to Rethink Teaching and Learning." 9 jun. 2026. https://tihupe.ee/en/3385/; Ministério da Educação e Pesquisa da Estônia. "Estonia Announces A Groundbreaking National Initiative: AI Leap Programme to Bring AI Tools to All Schools." https://www.hm.ee/en/news/estonia-announces-groundbreaking-national-initiative-ai-leap-programme-bring-ai-tools-all

9. UNESCO. "AI competency framework for students" e "AI competency framework for teachers." Set. 2024. https://www.unesco.org/en/articles/ai-competency-framework-students; https://www.unesco.org/en/articles/ai-competency-framework-teachers

10. OCDE/União Europeia. "Empowering Learners for the Age of AI: An AI Literacy Framework for Primary and Secondary Education." 18 jun. 2026. https://www.oecd.org/en/publications/empowering-learners-for-the-age-of-ai_65cd27d4-en.html

11. Folha de S.Paulo. "Conselho veta uso de IA sem supervisão para alunos até o 5º ano do fundamental." 1º set. 2026. https://www1.folha.uol.com.br/educacao/2026/09/conselho-veta-uso-de-ia-sem-supervisao-para-alunos-ate-o-5o-ano-do-fundamental.shtml; Ferreira, Paula. "Conselho do MEC define nível de risco do uso de IA para

cada atividade na educação; veja cada um." O Estado de S. Paulo, 1º set. 2026. https://www.estadao.com.br/educacao/conselho-do-mec-define-nivel-de-risco-do-uso-de-ia-para-cada-atividade-na-educacao-veja-cada-um/

CAPÍTULO 24

Como contratar quando todos podem parecer excelentes no papel

Tese do capítulo: currículos, cartas e testes para casa podem ser produzidos por IA; contratar exige observar pensamento, entrega, julgamento, honestidade e capacidade de aprender com IA.

Terminei o capítulo anterior dizendo que as empresas enfrentam o mesmo problema da escola. Se a lição de casa já não prova quem aprendeu, o currículo já não prova quem sabe fazer.

Passei décadas contratando do jeito tradicional. Ao longo da carreira, em grandes empresas e nas empresas que construí, o ritual era quase sempre o mesmo: uma pilha de currículos, uma triagem pelo diploma e pelas empresas anteriores, algumas entrevistas e uma boa dose de intuição. Esse ritual tinha falhas, mas partia de uma premissa razoável. Escrever um bom currículo dava trabalho, e esse trabalho dizia alguma coisa sobre a pessoa.

Essa premissa acabou. Hoje, a pergunta que eu me faço é outra: como descobrir quem realmente consegue produzir, aprender e responder pelo resultado?

Já estive dos dois lados da mesa de entrevista. Houve uma fase da minha vida em que fiquei sem trabalho, e foram mais de 50 entrevistas até voltar. Em cerca de 40% delas, o entrevistador não

tinha lido o meu currículo. Do outro lado, prestei consultoria a uma grande empresa de tecnologia que abria uma fábrica de software no interior de São Paulo: foram 1.769 candidatos para 40 vagas, e os testes aplicados davam resultados muito parecidos. Desde então, confio na intuição, olho o candidato nos olhos e pergunto sobre situações práticas.

Os números mostram o tamanho da mudança. Em junho de 2025, o New York Times relatou que o LinkedIn recebia cerca de 11 mil candidaturas por minuto, mais de 45% acima do ano anterior, com a ajuda de ferramentas de IA que adaptam currículos e até enviam candidaturas automaticamente.[1] No Brasil, numa pesquisa da consultoria Robert Half feita em junho de 2026 com 500 responsáveis por contratação, 53% disseram ter problemas com candidaturas produzidas com IA. Entre eles, 74% afirmaram que o tempo para preencher uma vaga aumentou. E um terço das empresas acrescentou entrevistas ou etapas extras para confirmar o que os candidatos diziam saber.[2]

1. O fim do currículo e do diploma como provas suficientes de competência

O currículo nunca foi uma prova. Sempre foi um indício. Mas, quando escrevê-lo exigia esforço, o indício valia alguma coisa: mostrava organização, capacidade de escrita e algum cuidado. Hoje, qualquer pessoa pode produzir em minutos um currículo impecável, ajustado palavra por palavra ao anúncio da vaga. Quando todos parecem excelentes no papel, o papel deixa de distinguir.

Em abril de 2026, a mesma Robert Half informou que 66% dos gestores de contratação no Brasil acreditavam que a IA generativa aumentava os currículos falsos ou exagerados, e que 70% ainda

não tinham adotado nenhuma medida para lidar com o risco.[3] O problema vai além do exagero. Numa pesquisa da Gartner com 3.000 candidatos, feita no segundo trimestre de 2025, 6% admitiram ter participado de fraude em entrevista, fingindo ser outra pessoa ou deixando alguém fingir ser elas. A Gartner prevê que, até 2028, um em cada quatro perfis de candidatos no mundo será falso.[4] É uma previsão, não um fato. Mas é uma previsão que nenhum empresário deveria ignorar.

O diploma também perdeu parte do seu poder de sinal, como vimos no Capítulo 21. Muitas empresas anunciaram que passariam a contratar por competências, e não por diploma. Na prática, a mudança foi bem menor do que os anúncios. Um estudo do Burning Glass Institute com a Harvard Business School, publicado em fevereiro de 2024, estimou que, em 2023, a contratação baseada em competências tinha mudado o resultado de menos de uma em cada 700 contratações nos Estados Unidos.[5] Tirar a exigência do diploma do anúncio é fácil. Difícil é criar outra forma confiável de avaliar.

E a necessidade dessa outra forma é reconhecida pelos próprios recrutadores. Na pesquisa Future of Recruiting 2025, do LinkedIn, feita em setembro de 2024 com 1.271 profissionais de recrutamento em 23 países, 93% disseram que avaliar com precisão as competências dos candidatos é crucial para melhorar a qualidade das contratações. Os dados da própria plataforma indicaram que as empresas que mais buscam candidatos por competências têm 12% mais chance de fazer uma contratação de qualidade do que as que não fazem esse tipo de busca. Vale lembrar que é uma plataforma comercial medindo o próprio produto, com a sua própria definição de qualidade. É um sinal, não uma lei.[6]

NÃO CONFUNDA

Currículo é o que a pessoa diz que fez. *Credencial* é o sinal formal de que ela passou por um caminho. *Competência* é o que ela consegue fazer agora, diante de você.

Remover a exigência de diploma não é o mesmo que contratar por competências. Sem uma forma rigorosa de avaliar, a empresa só troca um atalho por outro.

2. As cinco capacidades: domínio, aprendizagem, fluência em IA, verificação e comunicação humana

Se o papel não basta, o que procurar? Eu organizo a resposta em cinco capacidades.

Domínio. Conhecimento real da área. No Capítulo 20, vimos o experimento com consultores do Boston Consulting Group: a IA ajudou muito nas tarefas dentro da sua capacidade e reduziu a chance de acerto numa tarefa fora dela. Só quem domina o assunto sabe de que lado da fronteira está.

Aprendizagem. A velocidade com que a pessoa aprende algo novo. O relatório Future of Jobs 2025, do Fórum Econômico Mundial, baseado numa pesquisa com mais de mil grandes empregadores de 55 economias, estimou que 39% das competências essenciais dos trabalhadores vão mudar até 2030.[7] Quem contrata para o cargo de hoje pode estar contratando para um cargo que vai deixar de existir.

Fluência em IA. Não é saber o nome das ferramentas. É saber quando usá-las, para quê, com quais dados e com quais limites. Inclui proteger informações da empresa e dos clientes.

Verificação. A capacidade de desconfiar da própria ferramenta, conferir números, checar fontes e perceber quando a resposta está bonita demais para ser verdade.

Comunicação humana. Explicar uma escolha, convencer uma equipe, dar uma má notícia a um cliente, ouvir. A IA pode redigir o e-mail. Não pode sustentar a conversa difícil que vem depois.

O currículo diz onde a pessoa esteve. A tarefa mostra o que ela consegue fazer.

3. O novo processo: entrevista estruturada, tarefa real com IA permitida, revisão crítica, referências e plano de 30-60-90 dias

Agora a parte prática. É assim que eu montaria um processo de contratação hoje.

Entrevista estruturada. As mesmas perguntas, na mesma ordem, para todos os candidatos, avaliadas com critérios definidos antes. Parece burocrático, mas é o método que a pesquisa mais apoia. Uma revisão de décadas de estudos sobre seleção de pessoal, publicada em 2022 no Journal of Applied Psychology por Paul Sackett e colegas, colocou a entrevista estruturada no topo da lista dos métodos que melhor preveem o desempenho no trabalho. Os autores alertam que o resultado varia bastante de uma empresa para outra, e que todos os métodos preveem menos do que se imaginava.[8] Nenhum método é infalível. Mas improvisar é pior.

Tarefa real com IA permitida. Em vez de proibir a IA, que o candidato vai usar no trabalho de qualquer jeito, entregue um problema real, com informação faltando, e permita o uso das ferramentas. Algumas grandes empresas já seguem esse caminho.

Em julho de 2025, segundo o site 404 Media, com apoio da revista Wired, a Meta informou aos funcionários que passaria a permitir que alguns candidatos a vagas de programação usassem um assistente de IA durante a entrevista. Segundo uma mensagem interna citada pela reportagem, o formato é "mais representativo do ambiente" em que os futuros funcionários vão trabalhar.[9] A tarefa deve ser curta, relevante e ética. Não pode virar trabalho gratuito disfarçado de teste.

Revisão crítica. Depois da tarefa, a conversa. O que você perguntou à IA? O que aceitou, o que rejeitou e por quê? Como verificou os números? O que faria diferente? Essa conversa é a defesa oral de que falei no Capítulo 23, agora na empresa. É nela que aparecem o julgamento e a honestidade.

Referências. Num mundo em que documentos podem ser fabricados, a conversa com quem trabalhou com a pessoa volta a valer muito. Pergunte por fatos e exemplos, não por adjetivos.

Plano de 30-60-90 dias. Peça ao finalista um plano para os primeiros três meses. Ele mostra se a pessoa entendeu o problema da empresa, sabe priorizar e tem noção de prazo. E vira o primeiro acordo de expectativas depois da contratação.

E a IA do lado de quem contrata? Ela pode ajudar a organizar candidaturas, resumir entrevistas e padronizar critérios. Mas automatizar decisões sobre pessoas tem riscos sérios. Nos Estados Unidos, o caso Mobley contra Workday, empresa de software de recursos humanos, acusa as ferramentas de triagem automática da empresa de discriminar candidatos por idade, raça e deficiência. A Workday nega. Em maio de 2025, a Justiça autorizou, de forma provisória, uma ação coletiva em nome de candidatos com 40 anos ou mais, e milhares aderiram. Até a data desta fotografia, o processo continua em andamento, sem decisão sobre o mérito.[10]

No Capítulo 12, vimos que a Europa classificou os sistemas de IA usados em seleção de pessoal como de alto risco, embora tenha adiado essas regras para dezembro de 2027. E os candidatos também desconfiam: na pesquisa da Gartner, só 26% confiavam que a IA os avaliaria de forma justa.[4]

FOTOGRAFIA DA FRONTEIRA — CONTRATAÇÃO E IA (6 DE OUTUBRO DE 2026)

Volume: cerca de 11 mil candidaturas por minuto no LinkedIn (jun. 2025).

Brasil (Robert Half, 2026): 66% dos gestores veem mais currículos falsos ou exagerados com IA; 53% relatam problemas; 33% acrescentaram etapas de avaliação.

Fraude: 6% dos candidatos admitiram fraude em entrevista; previsão da Gartner de um em cada quatro perfis falsos até 2028.

Competências: 39% das competências essenciais devem mudar até 2030 (Fórum Econômico Mundial, 2025).

Método: entrevista estruturada no topo dos métodos de seleção (Sackett e colegas, 2022).

Risco jurídico: Mobley contra Workday em andamento nos EUA; alto risco do AI Act adiado para dez. 2027.

Fatos vivos: confira o Atlas Vivo da IA.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: no futuro, ninguém vai acreditar só no que você escreve sobre si mesmo. Vão pedir para você mostrar. Comece agora a guardar provas do que você faz: projetos, trabalhos, coisas que você construiu e consegue explicar.

Para quem tem 65 anos: é como na época em que se contratava um marceneiro pedindo para ver uma peça pronta e conversando com quem já tinha trabalhado com ele. A carta de apresentação perdeu valor. A peça e a palavra de quem conhece o trabalho voltaram a valer.

O QUE AINDA NÃO SABEMOS

Se as tarefas com IA permitida preveem o desempenho melhor do que os testes tradicionais. Como os tribunais vão dividir a responsabilidade entre a empresa que contrata e o fornecedor do software de triagem. Se a fraude de identidade vai obrigar a volta das entrevistas presenciais. E como avaliar com justiça quem nunca teve acesso às melhores ferramentas.

O que mudou / O que significa / O que fazer agora

O que mudou: a IA tornou fácil produzir currículos, cartas e testes impecáveis, e multiplicou o volume de candidaturas. O papel deixou de distinguir quem sabe de quem parece saber.

O que significa: contratar volta a depender de observar a pessoa trabalhando, pensando e explicando. O processo fica mais trabalhoso, mas também mais justo para quem tem competência e não tinha o currículo certo.

O que fazer agora: na próxima vaga que você abrir, defina antes as perguntas e os critérios da entrevista. Inclua uma tarefa curta e real, com IA permitida, seguida de uma conversa sobre como a pessoa chegou ao resultado. Se usar IA na triagem, mantenha uma pessoa revisando as decisões. Não contrate quem sabe repetir o nome da ferramenta; contrate quem consegue definir o problema, produzir resultado e responder pelo que entrega.

AÇÃO EM DEZ MINUTOS

Pegue o anúncio de uma vaga da sua empresa, ou da sua própria função. No seu Diário de Fronteira, escreva uma tarefa real de 30 minutos que mostraria se alguém sabe fazer aquele trabalho, e três perguntas para a conversa depois da tarefa. Coloque a data.

Repare que, em quase todas as etapas desse processo, aparece a mesma habilidade: transformar um problema em instruções claras, para uma IA ou para uma equipe. Muita gente chama isso de "saber fazer prompt" e procura frases mágicas. No próximo capítulo, vamos ver por que prompt não é magia.

Notas

1. The New York Times (DealBook). "A.I. Sludge Has Entered the Job Search." 21 jun. 2025. https://www.nytimes.com/2025/06/21/business/dealbook/ai-job-applications.html; Edwards, Benj. "The résumé is dying, and AI is holding the smoking gun." Ars Technica, 24 jun. 2025. https://arstechnica.com/ai/2025/06/the-resume-is-dying-and-ai-is-holding-the-smoking-gun/

2. Época Negócios. "Candidaturas com IA já criam problemas para a maioria dos empregadores no Brasil, aponta pesquisa." 31 ago. 2026 (pesquisa da Robert Half com 500 responsáveis por contratação, jun. 2026). https://epocanegocios.globo.com/futuro-do-trabalho/noticia/2026/08/

candidaturas-com-ia-ja-criam-problemas-para-a-maioria-dos-empregadores-no-brasil-aponta-pesquisa.ghtml

3. Robert Half Brasil. "66% dos gestores de contratação afirmam que a IA generativa aumenta currículos falsos ou exagerados." Abr. 2026. https://www.roberthalf.com/br/pt/sobre-robert-half/imprensa/gestoresafirmam

4. Gartner. "Gartner Survey Shows Just 26% of Job Applicants Trust AI Will Fairly Evaluate Them." 31 jul. 2025. https://www.gartner.com/en/newsroom/press-releases/2025-07-31-gartner-survey-shows-just-26-percent-of-job-applicants-trust-ai-will-fairly-evaluate-them

5. Sigelman, Matthew; Fuller, Joseph B.; Martin, Alex. "Skills-Based Hiring: The Long Road from Pronouncements to Practice." Burning Glass Institute e Harvard Business School, fev. 2024. https://www.hbs.edu/managing-the-future-of-work/Documents/research/Skills-Based%20Hiring.pdf

6. LinkedIn. "The Future of Recruiting 2025" (pesquisa com 1.271 profissionais de recrutamento em 23 países, set. 2024, e dados da plataforma). https://business.linkedin.com/hire/resources/future-of-recruiting; LinkedIn Talent Blog. "LinkedIn Report: How AI Will Redefine Recruiting in 2025." https://www.linkedin.com/business/talent/blog/talent-acquisition/future-of-recruiting-2025

7. World Economic Forum. "The Future of Jobs Report 2025." Jan. 2025 (pesquisa com mais de 1.000 empregadores em 55 economias, mai.–set. 2024). https://www.weforum.org/publications/the-future-of-jobs-report-2025/digest/

8. Sackett, Paul R.; Zhang, Charlene; Berry, Christopher M.; Lievens, Filip. "Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range." Journal of Applied Psychology 107(11), 2022 (publicado on-line em 30 dez. 2021). https://doi.org/10.1037/apl0000994

9. Koebler, Jason. "Meta Is Going to Let Job Candidates Use AI During Coding Tests." 404 Media (com apoio da Wired), 29 jul. 2025. https://www.404media.co/meta-is-going-to-let-job-candidates-use-ai-during-coding-tests/

10. Tribunal Distrital dos EUA para o Norte da Califórnia. Mobley v. Workday, Inc., n.º 3:23-cv-00770-RFL. Aviso sobre a ação coletiva provisoriamente certificada (out. 2025) e decisão de 2 jul. 2026 que negou recurso antecipado à Workday. https://storage.courtlistener.com/recap/gov.uscourts.cand.408645/gov.uscourts.cand.408645.177.0_1.pdf; https://case-law.vlex.com/vid/mobley-v-workday-inc-1127795073

CAPÍTULO 25

Prompt não é magia

Tese do capítulo: um bom prompt é pensamento claro transformado em direção; a habilidade durável não é conhecer uma frase secreta, mas definir objetivo, contexto, restrições, formato e critério de sucesso.

Terminei o capítulo anterior dizendo que, em quase todas as etapas de uma boa contratação, aparece a mesma habilidade: transformar um problema em instruções claras. Muita gente chama isso de "saber fazer prompt". E muita gente vende isso como mágica.

Você já deve ter visto as listas: "os 50 prompts que vão mudar sua vida", "a frase secreta que faz o ChatGPT trabalhar dez vezes melhor". Eu entendo o apelo. Quando uma tecnologia parece misteriosa, a gente procura um feitiço.

Mas a minha experiência mostra outra coisa. Muitas das minhas ideias nascem faladas, de forma espontânea, quase desorganizadas, do jeito que a cabeça explode. O trabalho de verdade começava depois, quando eu precisava transformar aquela ideia numa instrução que alguém conseguisse executar. E percebi que esse alguém podia ser uma inteligência artificial ou uma equipe. O esforço era exatamente o mesmo.

As histórias deste livro nasceram exatamente assim. Muitas eu ditei em português, no meu ritmo, do jeito que vinham à cabeça: "Mantena, Minas Gerais, Brasil, só vi um telefone quando tinha 11

anos, TV preta e branca...". Para virar capítulo, cada lembrança precisou de contexto: onde, quando, quem e por quê.

Prompt é a instrução, ou o conjunto de informações, que orienta a IA. E um prompt ruim quase nunca é um problema de vocabulário. É um problema de pensamento incompleto.

1. A anatomia do pedido: o Método CLARO

Todo líder conhece a cena. Você diz a alguém "cuida disso pra mim" e, dias depois, recebe algo completamente diferente do que imaginava. A culpa não é só de quem executou. A ambiguidade estava no pedido.

Com a IA é igual. A Anthropic, criadora do Claude, sugere nos seus guias oficiais tratar o modelo como um funcionário novo e brilhante, que não conhece o contexto da sua empresa. E propõe uma "regra de ouro": mostre o seu pedido a um colega com pouco contexto sobre o assunto. Se ele ficaria confuso, a IA também ficará.[1] No Capítulo 3, eu já tinha sugerido tratar um agente como um funcionário novo e muito rápido. A mesma lógica vale para qualquer pedido.

Para organizar isso, eu uso o que chamo de Método CLARO:

1. **C de Contexto.** A situação, o público, os materiais disponíveis e o que a IA precisa saber para não adivinhar.
2. **L de Limites.** Prazo, extensão, fontes permitidas, dados que não podem ser inventados, informações sensíveis e o que está fora do escopo.
3. **A de Ação.** O que exatamente precisa ser feito: analisar, comparar, calcular, escrever, revisar, planejar.
4. **R de Resultado.** O formato, a ordem, o tom, o nível de detalhe e o critério que vai tornar a resposta utilizável.

5. **O de Observação.** Pedir que a IA aponte premissas, incertezas e fontes, critique a própria resposta, e depois refinar o pedido com base no que voltou.

O método não é invenção minha no conteúdo. Ele resume o que os próprios fabricantes recomendam. O guia da Anthropic começa por exigir critérios de sucesso definidos e uma forma de testá-los antes de qualquer ajuste no prompt.[1] O Google, nos guias do Gemini, recomenda instruções claras e específicas, contexto suficiente, exemplos com formato consistente e divisão de tarefas complexas em etapas.[2] A OpenAI, nas orientações para os seus modelos de raciocínio, pede objetivo bem definido, critérios de sucesso específicos e separação clara entre as partes do pedido.[3] Três empresas concorrentes, a mesma lição: clareza vence truque.

No Apêndice C, deixei um laboratório completo do Método CLARO, com um modelo pronto para copiar e exemplos de pedidos fracos transformados em pedidos executáveis.

NÃO CONFUNDA

Prompt é o pedido que você escreve.

Contexto é tudo o que a IA recebe junto com ele: documentos, histórico da conversa, instruções da ferramenta. Em setembro de 2025, a Anthropic passou a chamar o cuidado com esse conjunto de "engenharia de contexto", e a descreveu como a evolução natural da engenharia de prompt.[4]

Critério de sucesso é como você vai saber se a resposta presta. Sem ele, nenhum prompt é bom ou ruim. É só diferente.

2. O que fazer: fornecer materiais, exemplos e critérios; dividir tarefas; pedir fontes; iterar e testar

Forneça materiais. A IA trabalha muito melhor com o seu documento, a sua planilha, o seu exemplo do que com a memória genérica dela. O guia do Google recomenda incluir o contexto relevante em vez de supor que o modelo já sabe.[2] Mas só compartilhe o que pode ser compartilhado com segurança, como veremos adiante.

Dê exemplos. Um bom exemplo do resultado esperado vale mais do que três parágrafos de explicação. Os três fabricantes recomendam exemplos, com uma ressalva da OpenAI para os modelos de raciocínio: tente primeiro sem exemplos e só acrescente se precisar, cuidando para que eles combinem com as instruções.[3]

Defina critérios. Diga como será usada a resposta, para quem e o que a torna boa.

Divida tarefas. Um pedido gigante, com dez objetivos ao mesmo tempo, produz uma resposta média em tudo. Peça primeiro a estrutura, depois cada parte, depois a revisão.

Peça fontes, e abra as fontes. No Capítulo 3, vimos que a IA pode inventar com total segurança, e no Capítulo 22 vimos referências inventadas em artigos aprovados por especialistas. Pedir fontes não basta. É preciso conferir.

Itere e teste. O primeiro pedido raramente é o melhor. A mesma pergunta pode receber respostas diferentes. Pesquisadores da Wharton School, da Universidade da Pensilvânia, mostraram isso em março de 2025: o mesmo modelo, com a mesma pergunta e o mesmo prompt, nem sempre acertava. Por isso, um pedido que vai

virar processo deve ser testado várias vezes antes de ganhar a sua confiança.[5]

Prompt fraco quase sempre é pensamento incompleto.

3. O que não fazer: confiar em um megaprompt, expor dados, misturar instruções ou aceitar a primeira resposta como verdade

Não acredite em frases mágicas. A mesma equipe da Wharton testou, ao longo de 2025, alguns dos truques mais repetidos na internet. Ser educado com a IA ou dar ordens a ela mudava o resultado de perguntas isoladas, para melhor ou para pior, mas o efeito quase desaparecia no conjunto.[5] Ameaçar o modelo ou oferecer uma gorjeta não trouxe melhora significativa no geral.[6] E dizer à IA que ela é "o maior especialista do mundo" em física ou em direito não melhorou a precisão das respostas em provas difíceis. Personas de pouco conhecimento, como "uma criança", chegaram a piorar o desempenho.[7] Os próprios autores lembram que a persona pode servir para outras coisas, como ajustar o tom. Mas não substitui dados, escopo e conhecimento real.

Não confie num megaprompt. Um texto enorme, com instruções acumuladas ao longo de meses, que ninguém consegue revisar, costuma esconder contradições. E não misture instruções: separe com clareza o que é contexto, o que é tarefa e o que é material. A OpenAI e o Google recomendam usar marcadores e títulos para delimitar cada parte.[2,3]

Não force o modelo a pensar como você acha que ele pensa. O conselho de pedir à IA que "pense passo a passo" virou mantra. Mas a OpenAI recomenda não fazer isso com os seus modelos de raciocínio, porque eles já raciocinam internamente, e a instrução pode até atrapalhar.[3] Em junho de 2025, a equipe da Wharton confirmou: nos modelos com raciocínio próprio, o ganho foi mínimo e o tempo de resposta aumentou de 20% a 80%.[8] Os modelos são diferentes, e um truque que funcionou em um pode não funcionar no seguinte.

Não exponha dados. Em abril de 2023, funcionários da divisão de semicondutores da Samsung colaram no ChatGPT código-fonte confidencial e anotações de uma reunião interna. Em maio, a empresa proibiu temporariamente o uso de ferramentas de IA generativa nos dispositivos corporativos enquanto criava um ambiente seguro.[9] Isso aconteceu no começo dessa história, mas a lição continua atual: segredos comerciais, dados pessoais, informações médicas e contratos só devem entrar em ferramentas aprovadas, com política clara. E, como vimos no Capítulo 18, quando a IA lê páginas e documentos de terceiros, comandos maliciosos escondidos neles também podem entrar no pedido.

Não aceite a primeira resposta como verdade. Uma resposta bem escrita não é uma resposta correta. Em assuntos jurídicos, médicos, financeiros, de segurança ou de emprego, a saída da IA é rascunho, nunca decisão.

FOTOGRAFIA DA FRONTEIRA — COMO PEDIR À IA (6 DE OUTUBRO DE 2026)

Guias oficiais: Anthropic, Google e OpenAI convergem em clareza, contexto, exemplos, critérios de sucesso, divisão de tarefas e teste.

Modelos de raciocínio: a OpenAI recomenda pedidos simples e diretos e dispensa o "pense passo a passo".

Pesquisa (Wharton, 2025): educação, ameaças, gorjetas e personas de especialista não melhoram a precisão de forma consistente; o "passo a passo" ajuda pouco nos modelos de raciocínio.

Nova expressão: "engenharia de contexto" (Anthropic, set. 2025).

Fatos vivos: confira o Atlas Vivo da IA.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: não existe código secreto. Pedir bem à IA é como explicar um trabalho em grupo para um colega que faltou à aula: diga o que é, para quando, como deve ficar e como saber se ficou bom.

Para quem tem 65 anos: é como passar uma tarefa a um funcionário novo, competente, mas que nunca trabalhou na sua empresa. Quanto mais claro o pedido, melhor o resultado. E você continua conferindo antes de assinar.

O QUE AINDA NÃO SABEMOS

Quanto da habilidade de pedir bem continuará necessária quando os modelos fizerem mais perguntas antes de responder. Se as técnicas que funcionam hoje vão funcionar na próxima geração de modelos. E como as empresas vão guardar, revisar e proteger os pedidos que viraram parte dos seus processos.

O que mudou / O que significa / O que fazer agora

O que mudou: os modelos ficaram muito melhores em entender linguagem comum. Truques e frases mágicas perderam valor, e os próprios fabricantes passaram a recomendar o básico: clareza, contexto, exemplos, critérios e teste.

O que significa: a habilidade durável não é decorar fórmulas. É pensar com clareza, delegar bem e saber avaliar o que voltou. É a mesma habilidade que sempre separou bons líderes de líderes confusos.

O que fazer agora: antes do próximo pedido importante a uma IA, escreva as cinco partes do Método CLARO. Peça premissas e fontes. Teste o mesmo pedido mais de uma vez e compare as respostas. E nunca coloque numa ferramenta um dado que você não colocaria num e-mail para fora da empresa. Prompting se tornará alfabetização básica. A vantagem real será saber o que perguntar, reconhecer uma boa resposta e perceber quando ela está errada.

AÇÃO EM DEZ MINUTOS

Pegue um pedido que você fez recentemente a uma IA e que não deu bom resultado. No seu Diário de Fronteira, reescreva-o com as cinco partes do Método CLARO. Envie as duas versões à mesma ferramenta e anote a diferença. Coloque a data.

Fim da Parte V

Chegamos ao fim da quinta parte. Perguntamos se a faculdade ainda importa e descobrimos que a resposta depende da pessoa,

da profissão, do preço e das alternativas. Vimos que um doutorado vale quando o problema exige profundidade, não quando serve de abrigo. Olhamos para uma escola que ainda ensina para um mundo que terminou, e para uma contratação em que o papel deixou de distinguir quem sabe de quem parece saber. E, agora, vimos que pedir bem a uma IA não é magia, é pensamento claro.

Aprender, trabalhar e contratar de novo era o quinto passo. O sexto é olhar para quem está vivendo essa revolução, e para o fato de que ninguém a vive da mesma forma.

Porque uma pessoa de 15 anos e uma de 70 não estão apenas em fases diferentes da vida. Foram formadas por tecnologias diferentes, aprenderam a pesquisar, a conversar e a confiar de jeitos diferentes. E agora precisam atravessar a mesma mudança, muitas vezes na mesma casa e na mesma empresa.

Na Parte VI, vamos olhar para quatro gerações na mesma revolução. Começando por quem tem quinze, vinte, vinte e cinco e trinta anos, e por uma pergunta que me intriga cada vez mais: quem vai ensinar quem?

Notas

1. Anthropic. "Prompt engineering overview" e "Prompting best practices" (seção "Be clear and direct"). Claude Docs, consultado em 6 out. 2026. https://docs.claude.com/en/docs/build-with-claude/prompt-engineering/overview; https://docs.anthropic.com/en/docs/build-with-claude/prompt-engineering/be-clear-and-direct

2. Google AI for Developers. "Prompt design strategies." Gemini API, consultado em 6 out. 2026. https://ai.google.dev/gemini-api/docs/prompting-strategies

3. OpenAI. "Reasoning best practices." OpenAI API Docs, consultado em 6 out. 2026. https://platform.openai.com/docs/guides/reasoning-best-practices

4. Rajasekaran, Prithvi et al. "Effective context engineering for AI agents." Anthropic Engineering, 29 set. 2025. https://www.anthropic.com/engineering/effective-context-engineering-for-ai-agents

5. Meincke, Lennart; Mollick, Ethan; Mollick, Lilach; Shapiro, Dan. "Prompting Science Report 1: Prompt Engineering is Complicated and Contingent." Wharton Generative AI Labs, 4 mar. 2025. https://arxiv.org/abs/2503.04818

6. Meincke, Lennart et al. "Technical Report: I'll pay you or I'll kill you — but will you care?" Wharton Generative AI Labs, 1º ago. 2025. https://gail.wharton.upenn.edu/research-and-insights/techreport-threaten-or-tip/

7. Basil, Savir et al. "Prompting Science Report 4: Playing Pretend: Expert Personas Don't Improve Factual Accuracy." Wharton Generative AI Labs, dez. 2025. https://arxiv.org/abs/2512.05858

8. Meincke, Lennart et al. "Prompting Science Report 2: The Decreasing Value of Chain of Thought in Prompting." Wharton Generative AI Labs, 8 jun. 2025. https://arxiv.org/abs/2506.07142

9. Gurman, Mark. "Samsung Bans ChatGPT, Google Bard, Other Generative AI Use by Staff After Leak." Bloomberg, 2 maio 2023. https://www.bloomberg.com/news/articles/2023-05-02/samsung-bans-chatgpt-and-other-generative-ai-use-by-staff-after-leak; Dobberstein, Laura. "Samsung reportedly leaked its own secrets through ChatGPT." The Register, 6 abr. 2023. https://www.theregister.com/software/2023/04/06/samsung-reportedly-leaked-its-own-secrets-through-chatgpt/895536

PARTE VI

Quatro gerações na mesma revolução

CAPÍTULO 26

Quinze, vinte, vinte e cinco, trinta

Tese do capítulo: pessoas separadas por apenas cinco anos podem ter sido formadas por experiências tecnológicas profundamente diferentes.

Terminei a Parte V com uma pergunta que me intriga cada vez mais: quem vai ensinar quem?

No Capítulo 3, contei o pequeno choque de ver gente jovem perguntando direto a uma inteligência artificial, sem abrir o Google, sem palavras-chave, sem lista de links. Aquela cena não saiu da minha cabeça. Com o tempo, percebi que ela escondia uma pergunta maior. Se eles não pesquisam como a minha geração pesquisava, também não aprendem, não confiam e não começam a carreira do mesmo jeito. E, se é assim, quem deveria estar ensinando quem?

Antes de seguir, um cuidado. Eu poderia dividir este capítulo em "geração Z" e "millennials", como fazem tantas reportagens. Prefiro não fazer isso. Em maio de 2023, o Pew Research Center, um dos institutos que mais estudam o tema, publicou um alerta: as fronteiras entre gerações não são definidas cientificamente, os rótulos podem virar estereótipos e as diferenças costumam ser exageradas, enquanto as semelhanças ficam esquecidas.[1] Por isso, vou falar de idades. E de um marco muito concreto.

O ChatGPT foi lançado em 30 de novembro de 2022.[2] Em outubro de 2026, quem tem 15 anos tinha 11 naquele dia. Quem tem 20 tinha 16. Quem tem 25 tinha 21, estava terminando a faculdade. E quem tem 30 tinha 26, já trabalhando. Quatro idades separadas por apenas cinco anos, e quatro primeiros contatos completamente diferentes com a mesma tecnologia.

1. Quinze: uma inteligência conversacional antes do primeiro emprego

Para quem tem 15 anos, a IA não é uma novidade. É parte da paisagem, como a internet era para quem nasceu nos anos 2000.

Os números brasileiros confirmam. A pesquisa TIC Kids Online Brasil, do Cetic.br, investigou pela primeira vez em 2025 o uso de IA generativa por crianças e adolescentes. Entre os usuários de internet de 9 a 17 anos, 65% tinham usado IA generativa para pelo menos uma das finalidades pesquisadas. Na faixa de 15 a 17 anos, eram 75%. Nessa idade, 68% usavam a IA para estudar ou fazer pesquisas escolares, 60% para procurar informações, 32% para criar textos, imagens, vídeos ou códigos, e 12% para conversar sobre problemas pessoais ou emoções.[3] No Capítulo 1, vimos números parecidos entre adolescentes americanos.

O que mais me chama a atenção é a ordem dos acontecimentos. Muita gente da minha geração conheceu o trabalho antes de conhecer o computador pessoal. Quem tem 15 anos hoje vai conhecer uma inteligência que conversa, escreve e explica antes do primeiro emprego, antes do primeiro chefe, antes da primeira entrevista. Vai chegar ao mercado sabendo pedir coisas a uma máquina com uma naturalidade que muitos adultos ainda não têm.

Mas naturalidade não é o mesmo que critério. No Capítulo 23, vimos que terceirizar o esforço pode melhorar a tarefa e prejudicar o aprendizado. Um adolescente que usa a IA para tudo pode chegar aos 18 anos com uma ferramenta poderosa nas mãos e pouca prática em pensar sem ela. E os 12% que conversam com a IA sobre emoções merecem atenção especial. Vamos voltar a esse ponto no Capítulo 29.

NÃO CONFUNDA

Idade é quantos anos a pessoa tem. *Geração* é um rótulo para quem nasceu num mesmo período, útil como referência, mas impreciso.

Familiaridade é usar uma ferramenta com naturalidade. *Compreensão* é saber como ela funciona, onde erra e o que custa usá-la. A primeira vem com a idade certa. A segunda precisa ser ensinada.

2. Vinte e vinte e cinco: estudo, identidade e início de carreira com degraus de entrada mudando

Quem tem 20 anos viveu a chegada do ChatGPT no meio do ensino médio. Quem tem 25 a viveu no fim da faculdade. São as pessoas que estão, agora, estudando, se formando e tentando o primeiro emprego enquanto as regras mudam.

Na universidade, o uso já é quase universal. No Reino Unido, a pesquisa anual do Higher Education Policy Institute, publicada em março de 2026 com 1.054 estudantes de graduação, mostrou que 95% usavam IA de alguma forma e 94% a usavam em trabalhos avaliados. Em 2024, só 66% usavam IA de algum jeito. Mas apenas 36% se sentiam incentivados pela própria instituição a usá-la.[4]

Nos Estados Unidos, o relatório da plataforma de empregos Handshake sobre a turma que se formou em 2026, publicado em abril, chamou esses estudantes de a primeira leva de talentos que viveu toda a faculdade junto com a IA, do lançamento do ChatGPT à primeira busca por emprego. Entre eles, 85% usavam IA, e mais de um terço todos os dias. Ao mesmo tempo, o pessimismo com a carreira subiu de 46% para 62% em dois anos. E 58% disseram que vão precisar de habilidades mais fortes em IA para ter sucesso no trabalho, mas só 28% afirmaram que o curso tinha integrado a IA de forma significativa. Um formando resumiu assim: "A IA parece uma habilidade que você ensina a si mesmo por sua própria conta e risco. Na faculdade, meus professores eram fortemente contra. Agora, os empregadores esperam que você já domine".[5]

Essa frase me tocou. Ela descreve uma geração presa entre duas instituições que ainda não conversam: a escola, que proíbe, e a empresa, que exige. No Capítulo 20, vimos que os mais jovens são os que mais sentem o aperto no degrau de entrada, e no Capítulo 21, que muitos recém-formados estão subempregados. Agora acrescento o lado de dentro: o que eles sentem.

E aqui está um dado que contraria o senso comum. Pensamos que quem mais usa a IA é quem mais acredita nela. Não é bem assim. Numa pesquisa do Pew com adultos americanos, feita em fevereiro de 2026, 66% dos que tinham de 18 a 29 anos usavam chatbots, a maior proporção entre todas as idades. Mas foram também os mais pessimistas: 48% acreditavam que a IA terá efeito negativo sobre a sociedade, contra 39% dos que tinham de 30 a 49 anos e 37% dos que tinham 50 ou mais.[6]

Quem cresce com uma tecnologia não é ingênuo em relação a ela. Às vezes, é justamente quem enxerga primeiro o seu preço.

Quem cresce com uma tecnologia aprende a usá-la. Nem sempre aprende a julgá-la.

3. Trinta: a geração-ponte que lembra do mundo sem ChatGPT, mas construirá a carreira com agentes

Quem tem 30 anos hoje ocupa um lugar muito especial. Tinha cerca de 11 anos quando o iPhone foi apresentado, em janeiro de 2007, como vimos no Capítulo 16. Aprendeu a pesquisar no Google e cresceu junto com as redes sociais. E tinha 26 anos, com alguns anos de trabalho, quando o ChatGPT chegou.

Ou seja: lembra do mundo antes da IA conversacional, mas vai construir quase toda a carreira com ela. Vai trabalhar com os agentes de que falamos ao longo deste livro e provavelmente vai liderar equipes que misturam pessoas e sistemas automáticos. E, entre os adultos americanos de 30 a 49 anos, 61% já usavam chatbots em 2026, uma proporção muito próxima da dos mais jovens.[6]

Eu chamo essas pessoas de geração-ponte. Elas conhecem os dois lados. Entendem a linguagem de quem tem 15 anos e a de quem tem 60. Sabem como era trabalhar sem IA, o que permite perceber o que se ganha e o que se perde com ela. E estão, muitas vezes, exatamente no meio da hierarquia das empresas, entre os estagiários que chegam usando IA para tudo e os diretores que ainda desconfiam dela.

Então, quem ensina quem? A minha resposta é: todos ensinam alguma coisa. Quem tem 15 ou 20 anos pode ensinar velocidade, naturalidade e coragem de experimentar. Quem tem 25 pode ensinar o que é começar uma carreira quando as regras mudam no

meio do caminho. Quem tem 30 pode traduzir entre os mundos. E quem tem mais idade pode ensinar o que nenhuma ferramenta ensina sozinha: julgamento, contexto, relacionamento e as consequências que só aparecem com o tempo.

Meu filho, o Filipe, se parece comigo o suficiente para saber exatamente onde me desafiar. Ele coloca quase todos os meus projetos em xeque e me obriga a defender o plano, mas nunca me disse que eu não posso ou não devo fazer alguma coisa. E admira que eu esteja quase sempre à frente do meu tempo, mesmo quando isso dá mais trabalho para nós dois. Talvez seja assim que as gerações devam se ensinar: uma desafia o plano, a outra sustenta a coragem.

FOTOGRAFIA DA FRONTEIRA — IDADES DIANTE DA IA (6 DE OUTUBRO DE 2026)

Marco: ChatGPT lançado em 30 nov. 2022. Em out. 2026, quem tem 15 anos tinha 11 naquele dia; quem tem 30 tinha 26.

Brasil (TIC Kids Online 2025): 75% dos usuários de internet de 15 a 17 anos usaram IA generativa; 12% para conversar sobre emoções.

Universidade (Reino Unido, HEPI 2026): 95% dos estudantes usam IA; 94% em trabalhos avaliados; 36% se sentem incentivados pela instituição.

Formandos (EUA, Handshake 2026): 85% usam IA; pessimismo com a carreira de 46% para 62% em dois anos.

Adultos (EUA, Pew 2026): 66% dos que têm 18–29 anos usam chatbots, e 48% deles esperam efeito negativo da IA sobre a sociedade.

Fatos vivos: confira o Atlas Vivo da IA.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: você sabe usar a IA melhor do que muitos adultos. Mas usar bem não é o mesmo que saber quando não usar. Pergunte aos mais velhos como eles tomavam decisões sem ela. Você pode se surpreender com o que eles sabem.

Para quem tem 65 anos: o jovem da sua família não está "brincando no celular". Muitas vezes, está aprendendo, criando ou pedindo ajuda. Peça a ele que mostre o que faz. E, em troca, ensine o que a vida ensinou a você.

O QUE AINDA NÃO SABEMOS

Que efeito terá crescer com uma inteligência conversacional desde a infância sobre memória, escrita, autonomia e relações. Se o pessimismo dos mais jovens é uma reação passageira a um mercado difícil ou uma leitura antecipada do futuro. E se as escolas e as empresas vão aprender a conversar a tempo de não deixar uma geração ensinando a si mesma por conta e risco.

O que mudou / O que significa / O que fazer agora

O que mudou: em menos de quatro anos, a IA conversacional passou a fazer parte da vida de quase todos os adolescentes e universitários. Pessoas separadas por cinco anos tiveram primeiros contatos completamente diferentes com ela.

O que significa: a diferença entre as idades não é só de habilidade com a ferramenta. É de formação, de expectativa e de medo. E a familiaridade dos mais jovens não substitui a experiência dos mais velhos, nem o contrário.

O que fazer agora: se você é jovem, procure alguém mais experiente e pergunte como ele decide, não apenas o que ele sabe. Se você é mais velho, peça a um jovem que mostre como usa a IA, sem julgar. Se você lidera uma empresa, crie duplas entre pessoas de idades diferentes, em que cada uma ensine o que sabe. Juventude oferece familiaridade, não imunidade. Saber usar uma ferramenta não é compreender suas consequências.

AÇÃO EM DEZ MINUTOS

Converse com alguém pelo menos dez anos mais novo ou mais velho que você. Pergunte como essa pessoa usou a IA na última semana e conte como você usou. No seu Diário de Fronteira, anote uma coisa que você aprendeu e uma que você ensinou. Coloque a data.

E se a juventude oferece familiaridade, mas não imunidade, a experiência oferece o oposto: julgamento, mas não familiaridade. No próximo capítulo, vamos olhar para quem tem quarenta, cinquenta, sessenta e setenta anos. E para por que a experiência pode virar uma vantagem extraordinária, desde que não se espere demais para começar.

Notas

1. Pew Research Center. "5 things to keep in mind when you hear about Gen Z, Millennials, Boomers and other generations." 22 maio 2023. https://www.pewresearch.org/short-reads/2023/05/22/5-things-to-keep-in-mind-when-you-hear-about-gen-z-millennials-boomers-and-other-generations/

2. OpenAI. "Introducing ChatGPT." 30 nov. 2022. https://openai.com/index/chatgpt/

3. Cetic.br/NIC.br. "TIC Kids Online Brasil: 65% das crianças e dos adolescentes usam IA generativa para estudar, criar conteúdo e lidar com

emoções." 22 out. 2025. https://cetic.br/noticia/tic-kids-online-brasil-65-das-criancas-e-dos-adolescentes-usam-ia-generativa-para-estudar-criar-conteudo-e-lidar-com-emocoes/; Cetic.br/NIC.br. "TIC Kids Online Brasil 2025: resumo executivo." 2026. https://www.cetic.br/media/docs/publicacoes/2/pt-br/20260317181001/tic_kids_online_2025_resumo_executivo.pdf.pdf

4. Higher Education Policy Institute (HEPI). "Student Generative AI Survey 2026." Mar. 2026 (1.054 estudantes de graduação em tempo integral no Reino Unido). https://www.hepi.ac.uk/reports/student-generative-ai-survey-2026/

5. Handshake. "AI and the workforce ahead: What the Class of 2026 tells us about the future of the labor market." Abr. 2026. https://joinhandshake.com/research/economic-research/class-of-2026-ai-outlook/

6. Pew Research Center. "How Americans' opinions and use of AI differ by age." 17 jun. 2026 (pesquisa de 17 a 23 fev. 2026). https://www.pewresearch.org/internet/2026/06/17/how-opinions-and-use-of-ai-differ-by-age/; apêndice com tabelas detalhadas: https://www.pewresearch.org/internet/2026/06/17/americans-and-ai-appendix-detailed-chart-and-tables/

CAPÍTULO 27

Quarenta, cinquenta, sessenta, setenta

Tese do capítulo: experiência acumulada pode se tornar vantagem extraordinária quando combinada com IA, mas esperar demais aumenta o custo da adaptação.

No fim do capítulo anterior, eu disse que a juventude oferece familiaridade, mas não imunidade. A experiência oferece o contrário: julgamento, mas não familiaridade. E, para muita gente com quarenta, cinquenta, sessenta ou setenta anos, essa falta de familiaridade pesa mais do que deveria.

Ao longo da minha carreira, vi chegar o Microsoft Office, o computador pessoal nas grandes redes de varejo, os celulares da Nokia e da Motorola, o BlackBerry. Em cada uma dessas revoluções, houve, para muita gente, um momento de estranhamento. Uma tela nova, uma palavra nova, um jeito novo de trabalhar. E, em cada uma delas, a saída foi a mesma: curiosidade. Sentar, testar, errar, perguntar, testar de novo.

Com a inteligência artificial, não está sendo diferente. Depois de décadas de carreira, continuo aprendendo sistemas novos. A diferença é que, agora, o sistema também conversa comigo e me ensina enquanto eu aprendo.

A mais recente foi aprender a ditar para os meus agentes de IA. Na primeira tentativa, o ditado do computador estava configurado em inglês, e eu falando português: saiu uma frase sem pé nem

cabeça. Bastou trocar o idioma, e as minhas histórias começaram a sair no meu próprio ritmo.

Insegurança diante do novo eu conheço bem. Quando cheguei à faculdade de Análise de Sistemas, no Rio de Janeiro, fiquei muito impressionada com o MS-DOS, com os computadores IBM e com a internet que era só texto. O curso exigia inglês, e eu fiz inglês intensivo, duas horas por dia, durante um ano. Quando tudo passou para o Windows, a minha vida ficou mais fácil. Depois, virei gerente de produtos da Microsoft e fui beta tester de muitos softwares da Microsoft e, na época, também da Lotus e da Autodesk. Hoje, quando aprendo uma ferramenta nova depois dos cinquenta, lembro que isso não é mais difícil do que foi aprender inglês naquela época. Só parece mais difícil porque agora o conforto está disponível, e naquela época não estava.

Este capítulo é para quem tem mais estrada do que tem tempo para aprender tudo de novo. E a mensagem central é simples: a sua experiência não é um peso. É matéria-prima. Mas ela só vira vantagem se você começar.

1. A experiência como base de julgamento, relacionamento e contexto

Um dos estudos mais citados sobre IA no trabalho ajuda a entender o que a experiência vale nesse novo mundo. Os economistas Erik Brynjolfsson, Danielle Li e Lindsey Raymond acompanharam 5.172 atendentes de suporte ao cliente de uma empresa que passou a oferecer a eles um assistente de IA generativa. O resultado foi publicado em 2025 no *Quarterly Journal of Economics*. Em média, a produtividade, medida em problemas resolvidos por hora, aumentou 15%. Para os atendentes menos experientes e menos qualificados, o aumento foi de 30%. Já

os mais experientes e mais qualificados tiveram ganhos pequenos de velocidade e pequenas quedas de qualidade.[1]

À primeira vista, parece uma má notícia para quem tem experiência. Mas olhe de novo. Os autores encontraram indícios de que a IA espalhava, para os novatos, as boas práticas dos atendentes mais capazes. Em outras palavras, segundo os autores, boa parte do valor do assistente vinha do que ele tinha aprendido com o trabalho dos melhores.[1] A experiência dos veteranos não desapareceu. Ela foi parar dentro da ferramenta.

Isso muda a pergunta. Não é mais "a IA vai saber o que eu sei?". Em muitos casos, vai saber uma parte. A pergunta passa a ser: o que eu sei que nenhuma ferramenta sabe sozinha?

A minha resposta tem três partes.

A primeira é **julgamento**. Quem já viu um projeto dar errado reconhece os sinais antes. Sabe quando um número está bonito demais, quando um contrato tem uma cláusula estranha, quando uma solução elegante não vai funcionar com aquela equipe. A IA pode produzir dez opções em segundos. Escolher a certa continua exigindo alguém que já errou o bastante.

A segunda é **relacionamento**. Décadas de carreira constroem uma rede de pessoas que confiam em você, e em quem você confia. Nenhuma IA herda isso. Ela pode escrever o e-mail. Não pode ser a pessoa para quem o cliente atende o telefone.

A terceira é **contexto**. A experiência ensina por que as coisas são como são: a história daquele mercado, as regras não escritas daquela empresa, o que já foi tentado e por que não deu certo. Como vimos no Capítulo 25, a IA responde melhor quando recebe contexto. E quem tem mais contexto para dar é justamente quem viveu mais.

Quando coloquei a American Photonics à venda, o telefone tocou 79 vezes. A maioria dos interessados era educada, charmosa, fluente na língua das planilhas. Mas, quando eu escutava além do vocabulário, o plano era quase sempre o mesmo: comprar a empresa, demitir todo mundo, inflar a receita com um orçamento de fome e revendê-la, inteira ou em pedaços. Ninguém dizia isso desse jeito, claro. Mas o tamanho das perguntas revela a verdade, como o tamanho de uma aposta. Os compradores que perguntavam sobre as minhas máquinas e as minhas margens, mas nunca sobre as minhas pessoas, já tinham mostrado as cartas. Os que corriam para o preço nos primeiros cinco minutos também. Nenhuma planilha me mostrou isso. Trinta anos de pôquer mostraram. Eu tenho funcionários que entraram na empresa em 2001 e nunca saíram.

NÃO CONFUNDA

Experiência é o que você aprendeu vivendo. *Atualização* é o que você aprende sobre as ferramentas de agora. Uma não substitui a outra.

Lentidão é demorar por não saber. *Cuidado* é demorar porque você sabe o que pode dar errado. Quem tem mais estrada às vezes é confundido com lento quando está, na verdade, sendo cuidadoso.

2. O medo de parecer lento, velho ou desatualizado

Agora a parte difícil. A experiência vale muito, mas o mercado nem sempre a enxerga assim. Em maio de 2024, o Work Trend Index, pesquisa anual da Microsoft e do LinkedIn, mostrou que 71% dos líderes entrevistados disseram que prefeririam contratar um candidato menos experiente com habilidades em IA a um mais

experiente sem elas.[2] É um dado de dois anos atrás, e é uma pesquisa de uma empresa que vende IA. Mas ele mostra como o vento já soprava.

Esse medo tem fundamento, e os números mostram que a distância existe. No Brasil, a pesquisa TIC Domicílios 2025 perguntou pela primeira vez sobre o uso de IA generativa. Entre os usuários de internet de 16 a 24 anos, 55% tinham usado. De 35 a 44 anos, 30%. De 45 a 59, 18%. Com 60 anos ou mais, apenas 6%.[3] Nos Estados Unidos, a distância é menor, mas também aparece: na pesquisa do Pew que citei no capítulo anterior, 42% dos adultos de 50 a 64 anos e 23% dos que tinham 65 ou mais usavam chatbots.[4]

O dado mais revelador, porém, não é quem usa. É por que não usa. Entre os brasileiros com 60 anos ou mais que estão na internet e não usaram IA, 72% apontaram falta de habilidade para usar esse tipo de ferramenta, 68% citaram preocupações com segurança ou privacidade e 66% disseram não saber que esse tipo de ferramenta existia. Na faixa de 45 a 59 anos, 65% apontaram falta de habilidade e 71% citaram segurança ou privacidade.[5]

Ou seja: para muita gente mais velha, o problema não é ser contra. É não saber por onde começar, não ter quem mostre e ter medo de errar ou de se expor.

E a idade não explica tudo. Na mesma pesquisa brasileira, 59% dos usuários de internet com ensino superior tinham usado IA generativa, contra 17% dos que tinham ensino fundamental.[3] Renda, escolaridade e oportunidade pesam tanto quanto o ano de nascimento.

Há também sinais de movimento. Nos Estados Unidos, a pesquisa anual de tecnologia da AARP, organização dedicada a pessoas com 50 anos ou mais, mostrou que o uso de IA nessa faixa

quase dobrou, de 18% em 2024 para 30% em 2025. O mesmo levantamento registrou que três em cada cinco adultos com 50 anos ou mais sentem que a tecnologia não é projetada pensando na idade deles.[6]

Esse último dado diz muito. Parte do medo de parecer lento não está nas pessoas. Está em produtos feitos por jovens, para jovens, com uma linguagem que pressupõe que todo mundo já sabe o que é um "prompt" ou um "agente".

Já perdi a conta das vezes. Uma das que mais lembro foi numa mesa final em Sarasota, num torneio que começou com mais de 500 jogadores no One-Eyed Jacks. Um adversário, tentando me desestabilizar, perguntou em voz alta de onde eu era. Ele já sabia. Respondi, calma: "Do Brasil". Ele sorriu de lado: "Ah, minha faxineira também é do Brasil. Você limpa casas?". Respirei, sorri e respondi: "Não. Mas, se limpasse, você não teria dinheiro para me pagar". A mesa caiu na gargalhada, e quem ficou desconcertado foi ele. Fui a primeira da minha família a terminar a faculdade, uma das primeiras mulheres em cargos em que os homens achavam que mandavam, e uma imigrante aprendendo as regras não escritas de um país novo. A primeira vez raramente é a mais difícil. A mais difícil é a sétima.

A experiência só vira vantagem quando encontra a ferramenta. Esperando, ela vira saudade.

3. Como começar sem se tornar programador nem abandonar tudo o que já se sabe

A boa notícia é que nunca foi tão fácil começar. A IA de hoje entende linguagem comum. Você não precisa aprender uma

linguagem de programação para usá-la. Precisa aprender a conversar com ela, e isso, quem tem décadas de reuniões, negociações e briefings já sabe fazer melhor do que imagina.

E há uma notícia ainda melhor, que vem da ciência do envelhecimento. Num estudo publicado em 2014 na revista *Psychological Science*, pesquisadores da Universidade do Texas em Dallas, liderados pela psicóloga Denise Park, distribuíram adultos mais velhos em grupos diferentes durante três meses. Alguns aprenderam habilidades novas e exigentes, como fotografia digital ou costura de patchwork, por cerca de 15 a 16 horas por semana. Outros fizeram atividades sociais ou tarefas mentais simples. Só os que aprenderam algo novo e difícil tiveram melhora na memória.[7]

Não é uma prova de que a IA rejuvenesce ninguém. Mas é um lembrete de que aprender coisas novas e desafiadoras faz bem ao cérebro em qualquer idade. O desconforto do começo não é sinal de que você está velho demais. É sinal de que você está aprendendo.

Então, como começar? Cinco passos simples:

1. **Comece por um problema que você conhece bem.** Não pela ferramenta da moda. Um relatório que você faz todo mês, uma carta difícil, um contrato para revisar. Como você conhece o assunto, vai perceber na hora quando a IA errar. A sua experiência vira o seu controle de qualidade.
2. **Use a voz, se preferir.** A maioria das IAs principais aceita perguntas faladas. Fale como falaria com um assistente.
3. **Peça para a IA ensinar você.** "Me explique como se eu nunca tivesse usado isso." "Qual é a próxima coisa que eu deveria aprender?" Ela não se cansa de responder, não revira os olhos e não comenta com os colegas.

4. **Dê contexto, como você daria a um funcionário novo.** É o método CLARO do Capítulo 25 e do Apêndice C. Quem já treinou equipes já sabe fazer isso.
5. **Siga um ritmo, não uma maratona.** A Regra 1-1-1-30 do Capítulo 4 e o Plano de 30 dias do Apêndice A foram pensados exatamente para isso: uma IA principal, um processo real, uma revisão a cada 30 dias.

Imagine uma contadora de 58 anos, com trinta anos de profissão. Ela não precisa virar especialista em IA. Precisa descobrir que pode pedir a uma IA um primeiro rascunho da carta que explica a um cliente uma mudança tributária, e que é ela, com toda a sua experiência, quem vai perceber se a carta está certa, se o tom está adequado e se aquele cliente vai entender. A máquina traz velocidade. Ela traz julgamento, relacionamento e contexto.

E experiência não nos torna imunes ao medo. Sei disso por causa de um dia de pôquer. No Main Event da World Series of Poker de 2024, com mais de 10.000 jogadores, sobrevivi ao Dia 1. Depois, com a premiação se aproximando, as câmeras, os rostos famosos que eu tinha visto na televisão durante anos, alguma coisa dentro de mim simplesmente desligou. Deu um pane: o sistema travou. As minhas mãos gelaram sobre as fichas, as minhas leituras ficaram em silêncio, e eu desisti de mãos que eu daria bronca numa aluna por desistir. Não cheguei ao Dia 3. Medo do novo não é falta de capacidade. É um sistema que trava, e que a gente aprende a reiniciar. E, se um amigo da minha idade me dissesse "isso não é mais para mim", eu sei o que responderia.

Eu diria: você não está atrasado; você está em posição. No pôquer, posição significa agir com mais informação do que quem agiu antes. A idade é a mesma vantagem, disfarçada. Aprender algo novo depois dos cinquenta não é mais difícil do que foi, para mim, aprender inglês para a faculdade, duas horas por dia,

durante um ano. Só parece mais difícil porque agora o conforto está disponível. Recuse o conforto. Marque o aprendizado na agenda como uma reunião com o seu cliente mais importante: trinta minutos, uma pergunta, uma anotação com as suas próprias palavras. Quem diz que já sabe o suficiente está se oferecendo para virar o adversário favorito de alguém. Estar afiado não é algo que se guarda. É algo que se renova.

FOTOGRAFIA DA FRONTEIRA — EXPERIÊNCIA DIANTE DA IA (6 DE OUTUBRO DE 2026)

Brasil (TIC Domicílios 2025): usaram IA generativa 55% dos usuários de internet de 16 a 24 anos; 30% de 35 a 44; 18% de 45 a 59; 6% com 60 anos ou mais.

Por que não usam (60+, Brasil): falta de habilidade, 72%; segurança ou privacidade, 68%; não sabiam que existia, 66%.

EUA (Pew, fev. 2026): usam chatbots 42% dos adultos de 50 a 64 anos e 23% dos que têm 65 ou mais.

EUA (AARP): uso de IA entre pessoas com 50+ passou de 18% (2024) para 30% (2025).

Registro histórico (maio 2024): 71% dos líderes ouvidos pela Microsoft e pelo LinkedIn preferiam um candidato menos experiente com habilidades em IA.

Fatos vivos: confira o Atlas Vivo da IA.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: quando você ensina a sua avó ou o seu professor a usar a IA, não faça por eles. Mostre e deixe que façam. E aproveite para perguntar o que eles fariam com a resposta. Você vai aprender mais do que ensinar.

Para quem tem 65 anos: lembra quando aprendeu a dirigir? No começo, era preciso pensar em cada pedal. Depois de algumas semanas, você só pensava no caminho. Com a IA é parecido. As primeiras semanas são as mais difíceis, e elas passam.

O QUE AINDA NÃO SABEMOS

Se as empresas vão continuar preferindo habilidade com IA à experiência, ou se vão perceber que precisam das duas juntas. Se os ganhos dos mais experientes com IA vão crescer à medida que as ferramentas ficarem mais capazes em tarefas complexas e raras. E se os produtos de IA vão, finalmente, ser desenhados pensando também em quem tem mais de 50 anos.

O que mudou / O que significa / O que fazer agora

O que mudou: a IA aprendeu com o trabalho dos mais experientes e passou a oferecer parte desse conhecimento a qualquer iniciante. Ao mesmo tempo, o mercado começou a valorizar a habilidade com IA, e a distância entre quem usa e quem não usa ainda é grande entre os mais velhos.

O que significa: a experiência não perdeu valor, mas mudou de função. Ela deixa de ser só o que você sabe e passa a ser o que

permite julgar, corrigir e dar contexto ao que a máquina produz. E cada mês de espera torna o primeiro passo um pouco mais caro.

O que fazer agora: escolha um problema que você conhece bem e resolva-o com a ajuda de uma IA nesta semana. Use a sua experiência para revisar a resposta. Se precisar, peça ajuda a alguém mais jovem, e ofereça a sua em troca. Você não está velho demais. Mas o tempo de adiar indefinidamente terminou.

AÇÃO EM DEZ MINUTOS

Escreva uma tarefa do seu trabalho ou da sua vida que você conhece tão bem que perceberia qualquer erro. Peça a uma IA que faça um primeiro rascunho dela. Anote no seu Diário de Fronteira o que a IA acertou, o que errou e o que só você sabia. Coloque a data.

E há uma mudança que atravessa todas as idades, mas que cada geração sente de um jeito. Para muita gente com mais estrada, ligar era o jeito natural de resolver as coisas. Para muitos jovens, uma ligação inesperada parece quase uma invasão. No próximo capítulo, vamos falar de como a tecnologia mudou o significado de ligar, responder e esperar. E de por que, hoje, uma ligação já é o segundo nível de intimidade.

Notas

1. Brynjolfsson, Erik; Li, Danielle; Raymond, Lindsey. "Generative AI at Work." *The Quarterly Journal of Economics*, v. 140, n. 2, p. 889–942, maio 2025 (publicação antecipada em 4 fev. 2025). https://doi.org/10.1093/qje/qjae044; resumo: https://www.gsb.stanford.edu/faculty-research/publications/generative-ai-work

2. Microsoft e LinkedIn. "2024 Work Trend Index Annual Report: AI at Work Is Here. Now Comes the Hard Part." 8 maio 2024.

https://news.microsoft.com/source/2024/05/08/microsoft-and-linkedin-release-the-2024-work-trend-index-on-the-state-of-ai-at-work/

3. Cetic.br/NIC.br. TIC Domicílios 2025, tabela M1 – "Usuários de internet, por uso de ferramentas de inteligência artificial (IA) generativa." Dez. 2025. https://cetic.br/pt/tics/domicilios/2025/individuos/M1/

4. Pew Research Center. "How Americans' opinions and use of AI differ by age." 17 jun. 2026 (pesquisa de 17 a 23 fev. 2026). https://www.pewresearch.org/internet/2026/06/17/how-opinions-and-use-of-ai-differ-by-age/

5. Cetic.br/NIC.br. TIC Domicílios 2025, tabela M3 – "Usuários de internet que não usaram IA generativa, por motivo declarado para não ter usado IA generativa." Dez. 2025. https://cetic.br/pt/tics/domicilios/2025/individuos/M3/

6. Kakulla, Brittne. "2026 Tech Trends and Adults 50-Plus." AARP Research, dez. 2025 (página atualizada em 20 jul. 2026; pesquisa on-line com 3.838 adultos americanos, 9 set. a 6 out. 2025). https://doi.org/10.26419/res.01020.001; https://www.aarp.org/pri/topics/technology/internet-media-devices/2026-technology-trends-older-adults/

7. Park, Denise C. et al. "The Impact of Sustained Engagement on Cognitive Function in Older Adults: The Synapse Project." *Psychological Science*, v. 25, n. 1, p. 103–112, jan. 2014 (on-line em 8 nov. 2013). https://doi.org/10.1177/0956797613499592; https://pubmed.ncbi.nlm.nih.gov/24214244

CAPÍTULO 28

Uma ligação já é o segundo nível de intimidade

Tese do capítulo: a tecnologia não apenas muda o canal; muda o significado social de interromper, responder, esperar e estar disponível.

No fim do capítulo anterior, eu disse que há uma mudança que atravessa todas as idades, mas que cada geração sente de um jeito. Ela cabe numa frase que me escapou quase sem pensar, e que acabou virando o título deste capítulo: hoje, para muitas pessoas, uma ligação já é o segundo nível de intimidade.

No Prólogo, usei essa ideia para mostrar que a revolução da IA não está só no trabalho. Está também na forma como nos aproximamos uns dos outros. Agora quero olhar para ela com calma. Primeiro vem o texto. Depois, se houver confiança, vem a voz. E, cada vez mais, entre as duas pessoas, há uma inteligência artificial escrevendo, resumindo, traduzindo ou atendendo por uma delas.

1. Quando ligar era cuidado, eficiência e normalidade

Durante a maior parte da história do telefone, ligar não era invasão. Era o jeito normal de resolver as coisas. E, muitas vezes, era também uma forma de carinho.

O telefone fixo estava preso a um lugar. Quando você ligava, não ligava para uma pessoa. Ligava para uma casa, para um escritório, para uma loja. Quem atendesse atendia. Por isso, ligar exigia uma etiqueta própria: "Boa tarde, desculpe incomodar, a Maria está?". E a ligação tinha horário. Ninguém ligava tarde da noite, a não ser em caso de emergência. Como vimos no Capítulo 19, o interurbano era caro, e muitas famílias esperavam a noite ou o fim de semana para falar com quem estava longe.

Em 3 de abril de 1973, Martin Cooper, engenheiro da Motorola, fez nas ruas de Nova York a primeira ligação pública de um telefone celular portátil. Ligou justamente para o chefe do projeto rival, nos Laboratórios Bell.[1] O primeiro celular comercial portátil da Motorola, o DynaTAC 8000X, só seria aprovado nos Estados Unidos dez anos depois, em 1983.[1] A partir dali, ligar foi deixando de ser falar com um lugar e passou a ser falar com uma pessoa, onde quer que ela estivesse.

Em vendas, por muito tempo, o telefone foi ferramenta de trabalho essencial. Era por ele que se fechava um pedido, se acalmava um cliente, se resolvia um problema de entrega. Uma ligação mostrava interesse. Um retorno rápido mostrava respeito.

Na minha vida, o telefone também já foi o aparelho mais cruel do mundo. Naquela fase em que fiquei sem trabalho e fiz mais de cinquenta entrevistas, eu aprendi como é triste esperar o telefone tocar. O recrutador promete ligar na semana seguinte, não diz o dia, e você passa a carregar o telefone para todo lado, até para o banheiro, ligando para o próprio número para conferir se o toque está alto o suficiente.

Eu cresci em Mantena, no interior de Minas Gerais. Vi um telefone pela primeira vez aos 11 anos. Em casa, a televisão era em preto e branco.

Mesmo quando o texto chegou, a voz continuou mandando por muito tempo. Em 2011, uma pesquisa do Pew Research Center mostrou que, nos Estados Unidos, 53% dos donos de celular preferiam ser contatados por ligação, contra 31% que preferiam mensagem de texto. Mas os sinais da mudança já estavam lá: entre os jovens de 18 a 24 anos, a média era de 109,5 mensagens enviadas ou recebidas por dia.[2]

2. Quando o texto passou a proteger tempo e a ligação virou urgência, invasão ou intimidade

A primeira mensagem de texto da história foi enviada em 3 de dezembro de 1992. Um engenheiro britânico de 22 anos, Neil Papworth, mandou de um computador para o celular de um diretor da Vodafone duas palavras: "Merry Christmas", feliz Natal.[3] Ninguém imaginava que aquilo mudaria o significado do próprio telefone.

O texto trouxe três presentes. O primeiro foi o **tempo**: quem recebe responde quando pode. O segundo foi o **controle**: dá para pensar, escrever, apagar, reescrever. O terceiro foi a **proteção**: ninguém vê sua hesitação, ninguém ouve sua voz tremer.

E, com esses presentes, a ligação mudou de sentido. Em 2024, uma pesquisa da empresa britânica de comparação de preços Uswitch, com 2.000 pessoas, citada pela BBC, indicou que um quarto dos jovens de 18 a 34 anos nunca atendia o telefone, e que quase 70% deles preferiam uma mensagem a uma ligação. Mais da metade disse acreditar que uma ligação inesperada significa más notícias. A mesma reportagem ouviu a psicóloga Elena Touroni, que resumiu a diferença assim: as ligações "são mais expositivas e exigem um nível maior de intimidade", enquanto as mensagens permitem se conectar "sem se sentir vulnerável ou exposto".[4]

Ou seja: para muita gente jovem, a ligação deixou de ser o normal e passou a ser uma de três coisas. Pode ser **urgência**: alguém só liga se aconteceu algo grave. Pode ser **invasão**: alguém está exigindo a minha atenção agora, sem perguntar se posso. Ou pode ser **intimidade**: só ligo para quem eu tenho certeza de que quer me ouvir.

Os números mais recentes mostram que essa não é uma mania de jovens. É uma mudança de toda a sociedade, mais forte entre eles. Em fevereiro de 2026, o instituto YouGov perguntou a 2.442 adultos americanos como se comunicavam. Para 85%, as mensagens eram o meio usado várias vezes por semana; as ligações de voz vinham em segundo lugar, com 58%. E 68% disseram que as mensagens já tinham substituído pelo menos parte das ligações que faziam. Quando perguntados por que ainda ligavam, as principais respostas foram que a ligação parece mais pessoal ou emocionalmente próxima, ajuda a evitar mal-entendidos e serve melhor para conversas urgentes, sérias ou delicadas.[5]

É exatamente isso. A ligação não morreu. Ela foi promovida. Virou o canal das coisas importantes.

NÃO CONFUNDA

Não atender nem sempre é desinteresse. Para muitos, é uma forma de proteger o tempo e escolher quando responder.

Ligar nem sempre é invasão. Para muitos, é a forma mais clara de mostrar cuidado e resolver rápido.

O problema raramente é o canal. É quando cada lado acredita que a sua regra é a única.

3. Quando a IA escreve, resume, traduz e negocia dos dois lados da conversa

Agora entra o terceiro personagem. A inteligência artificial já não está só nos nossos computadores de trabalho. Está no telefone, nas mensagens, entre as pessoas.

Em junho de 2025, a Apple anunciou que o iPhone passaria a ter três recursos novos no aplicativo de telefone, lançados em setembro daquele ano. A Triagem de Chamadas atende números desconhecidos, pergunta quem está ligando e por quê, e só então deixa o dono decidir se atende. O Assistente de Espera fica na linha por você quando você está esperando um atendente. E a Tradução ao Vivo traduz a conversa enquanto ela acontece, com modelos que, segundo a empresa, rodam no próprio aparelho, inclusive em português do Brasil.[6] No início de 2025, o Google também testou nos Estados Unidos um recurso que liga para estabelecimentos comerciais em nome do usuário para perguntar preço e disponibilidade, avisando que a chamada é automatizada.[7]

Pense no que isso significa. A ligação, que já tinha virado o "segundo nível", agora pode ter um porteiro. Antes de você ouvir a voz de alguém, uma IA já perguntou quem é e o que quer. E, do outro lado, como vimos no Capítulo 17, já há empresas vendendo agentes de IA para o atendimento ao cliente de grandes companhias. Não é difícil imaginar uma ligação em que a IA de uma pessoa liga para a IA de uma empresa, as duas negociam um horário, e nenhum ser humano fala com outro.

Nas mensagens, a mudança é mais silenciosa. Na mesma pesquisa do YouGov, 67% dos americanos disseram não ter usado nenhum recurso de IA embutido em aplicativos de mensagem no mês anterior. Mas, entre os da geração Z, 22% já tinham usado a

IA para reescrever ou editar uma mensagem.[5] O começo é pequeno. A direção, não.

E o que acontece com a relação quando a máquina ajuda a escrever? Um estudo publicado em 2023 na revista *Scientific Reports*, liderado por Jess Hohenstein, da Universidade Cornell, testou isso em dois experimentos com mais de mil pessoas. Os participantes conversavam por mensagem, e alguns recebiam sugestões automáticas de resposta, as chamadas "respostas inteligentes". O resultado foi ambíguo, e por isso mesmo fascinante. Quem usava as sugestões conversava mais rápido e com linguagem mais positiva, e era avaliado pelo parceiro como mais cooperativo e próximo. Mas, quando o parceiro suspeitava que a outra pessoa estava usando IA, passava a avaliá-la de forma mais negativa.[8]

Ou seja: a IA pode melhorar a conversa. Mas, se descobrimos que ela estava ali, sentimos que algo foi tirado de nós.

No Prólogo, fiz algumas perguntas que continuam sem resposta. Se a IA escreveu o meu pedido de desculpas, ele continua sendo meu? Agora acrescento outra. Se a minha IA resumiu a sua mensagem, e a sua IA escreveu a resposta, o que exatamente nós dissemos um ao outro?

A ligação não morreu. Ela foi promovida a canal das coisas importantes.

FOTOGRAFIA DA FRONTEIRA — A CONVERSA MEDIADA (6 DE OUTUBRO DE 2026)

Registro histórico: primeira ligação pública de celular portátil, 3 abr. 1973 (Motorola); primeira mensagem de texto, 3 dez. 1992 ("Merry Christmas").

Registro histórico (2011, EUA, Pew): 53% dos donos de celular preferiam ligação; 31%, texto.

Registro histórico (2024, Reino Unido, Uswitch): um quarto dos jovens de 18 a 34 anos nunca atendia o telefone; quase 70% preferiam texto.

EUA (YouGov, fev. 2026): 85% usam mensagens várias vezes por semana; 58%, ligações; 68% dizem que as mensagens substituíram parte das ligações.

No telefone: triagem de chamadas, espera assistida e tradução ao vivo no iPhone desde set. 2025; teste do Google com IA que liga para empresas desde jan. 2025.

Fatos vivos: confira o Atlas Vivo da IA.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: para os seus avós, uma ligação era o jeito de dizer "eu me importo com você". Quando eles ligam, não estão invadindo. Estão fazendo carinho do jeito que aprenderam. Atenda de vez em quando.

Para quem tem 65 anos: para os seus netos, uma mensagem não é frieza. É educação: eles estão deixando você responder quando puder. Se quiser ligar, mande antes uma mensagem: "posso te ligar?". Você vai ver como a conversa melhora.

O QUE AINDA NÃO SABEMOS

Se, quando a IA estiver nos dois lados da conversa, as pessoas vão se entender melhor ou apenas mais rápido. Se vai surgir uma etiqueta para avisar quando uma mensagem foi escrita com ajuda da máquina. E se a voz humana, ao vivo, vai se tornar ainda mais rara ou, justamente por isso, ainda mais valiosa.

O que mudou / O que significa / O que fazer agora

O que mudou: a ligação deixou de ser o canal padrão e passou a ser o canal das coisas urgentes, sérias ou íntimas. E a IA começou a entrar no meio da conversa: atendendo, esperando, traduzindo, sugerindo e reescrevendo.

O que significa: quando o canal muda, muda também o significado de interromper, responder, esperar e estar disponível. E, quando a máquina escreve por nós, a autenticidade deixa de ser automática. Passa a ser uma escolha.

O que fazer agora: combine as regras com as pessoas importantes da sua vida e do seu trabalho: o que é para mensagem, o que é para ligação, o que é urgente. Use a IA para ajudar a escrever, mas releia e assuma cada palavra que enviar. E, nas conversas que mais importam, considere ligar. Quando duas IAs redigem a conversa, precisamos perguntar onde permaneceu a intenção humana.

AÇÃO EM DEZ MINUTOS

Escolha uma pessoa importante com quem você conversa quase só por mensagem. Ligue para ela, ou mande antes um "posso te ligar?". Depois, anote no seu Diário de Fronteira o que a voz trouxe que o texto não trazia. Coloque a data.

Se a IA já está entrando no meio das nossas ligações e mensagens, o próximo passo é ainda mais delicado. Para muitas pessoas, ela já não é só a intermediária da conversa. É a outra ponta. No próximo capítulo, vamos falar do que acontece quando a IA entra na amizade, no amor e na identidade.

Notas

1. Motorola Solutions. "Cell Phone Development" (história da empresa: demonstração do DynaTAC em Nova York, 3 abr. 1973; aprovação do DynaTAC 8000X pela FCC, 21 set. 1983). https://www.motorolasolutions.com/en_us/about/history/explore-motorola-heritage/cell-phone-development.html; Guinness World Records. "1973: First Mobile Phone Call." 2015. https://www.guinnessworldrecords.com/news/2015/8/60/1973-first-mobile-phone-call-392969

2. Smith, Aaron. "Americans and Text Messaging." Pew Research Center, 19 set. 2011. https://www.pewresearch.org/internet/2011/09/19/americans-and-text-messaging/

3. Vodafone. "25 years since the world's first text message." 4 dez. 2017. https://www.vodafone.com/news/newsroom/technology/25-anniversary-text-message

4. Rufo, Yasmin. "Why Gen Z & Millennials are hung up on answering the phone." BBC News, 26 ago. 2024 (pesquisa Uswitch com 2.000 pessoas). https://www.bbc.com/news/articles/crgklk3p70yo

5. Bali, Shinmin. "How Americans communicate in 2026: The rise of messaging & AI trends." YouGov, 26 fev. 2026 (pesquisa on-line com 2.442 adultos americanos, 3 a 5 fev. 2026). https://yougov.com/en-us/articles/54176-how-americans-communicate-in-2026-the-rise-of-messaging-ai-trends

6. Apple. "Apple elevates the iPhone experience with iOS 26." Comunicado, 9 jun. 2025. https://www.apple.com/newsroom/2025/06/apple-elevates-the-iphone-experience-with-ios-26/; Apple. "New versions of Apple's software platforms are available today." 15 set. 2025. https://www.apple.com/newsroom/2025/09/new-versions-of-apples-software-platforms-are-available-today/

7. TechCrunch. "Google's 'Ask for Me' feature calls businesses on your behalf to inquire about services, pricing." 30 jan. 2025. https://techcrunch.com/2025/01/30/googles-ask-for-me-feature-calls-businesses-on-your-behalf-to-inquire-about-services-pricing/

8. Hohenstein, Jess et al. "Artificial intelligence in communication impacts language and social relationships." *Scientific Reports*, v. 13, 2023. https://doi.org/10.1038/s41598-023-30938-9; https://pubmed.ncbi.nlm.nih.gov/37015964/

CAPÍTULO 29

Quando a IA entra na amizade, no amor e na identidade

Tese do capítulo: a IA está deixando de ser ferramenta de informação para ocupar papéis de professora, confidente, conselheira e companhia.

No fim do capítulo anterior, eu disse que, para muitas pessoas, a IA já não é só a intermediária da conversa. É a outra ponta. Não está ajudando a escrever para alguém. Está sendo o alguém.

No Prólogo, escrevi que escrevo também como mãe. E é como mãe que me faço uma pergunta que não me deixa em paz. O que vai acontecer quando uma criança consultar primeiro uma IA sobre o próprio corpo, sobre o primeiro amor, sobre um medo que não tem coragem de contar, sobre a carreira, sobre a família? Antes da mãe, antes do pai, antes da professora, antes do melhor amigo.

A minha mãe, a Preta, sempre foi a minha confidente e a minha melhor amiga. Era com ela que eu levava as perguntas mais difíceis da vida.

Talvez por isso um dos projetos mais importantes da minha vida tenha nascido de uma frase sobre ela. Um dia, eu acordei muito triste, com uma notícia triste, e me perguntei: do que eu preciso agora para não me sentir tão miserável e tão triste? A frase

que me veio à mente foi: *I need to call my mother*. Comprei o domínio na hora e comecei o projeto. Assim nasceu o I Need to Call My Mother: um livro com a vida da mãe contada na voz dela, em que um pequeno código QR na página abre a gravação. A história dela. A voz dela. Para sempre.

E ele é também uma forma de aprender. O projeto existe para entendermos como as pessoas mais velhas pensam e reagem, e isso vai nos ajudar a fazer humanoides mais humanos, com ética e com princípios. Hoje, ele é a base da proposta que apresentei à NSF, no programa SBIR, para um humanoide ético de companhia para idosos. Mas há uma linha que eu não cruzo: no I Need to Call My Mother, a IA não é a sua mãe e não fala por ela. Nós não clonamos vozes.

Este é o capítulo mais delicado do livro. Vou falar de solidão, de afeto e de saúde mental, com o cuidado que eu gostaria que as empresas de tecnologia tivessem.

1. Tutor, editor, terapeuta improvisado e amigo artificial

A IA entrou na vida das pessoas pela porta da utilidade. Primeiro, como tutora: explica a lição, corrige o exercício. Depois, como editora: melhora o texto, ajusta o tom. Mas ela não ficou parada nessa porta.

Já vimos sinais disso. No Capítulo 1, citei a pesquisa do Pew Research Center em que 12% dos adolescentes americanos disseram ter usado chatbots para apoio emocional ou aconselhamento. No Capítulo 26, vimos que, no Brasil, 12% dos usuários de internet de 15 a 17 anos tinham usado IA generativa para conversar sobre problemas pessoais ou emoções. Prometi voltar a esse ponto. É aqui.

Há também uma categoria inteira de produtos feitos para isso: os "companheiros de IA", aplicativos em que a pessoa conversa com um personagem que se apresenta como amigo, namorado, namorada ou confidente. Em julho de 2025, a organização americana Common Sense Media, que defende proteções para crianças no mundo digital, publicou uma pesquisa com 1.060 adolescentes de 13 a 17 anos. Setenta e dois por cento já tinham usado um companheiro de IA, e 52% usavam pelo menos algumas vezes por mês. Um terço usava esses aplicativos para interação social e relacionamentos: 12% para apoio emocional ou de saúde mental, 9% como amigo ou melhor amigo, 8% para interações românticas ou de flerte. E um terço dos adolescentes disse já ter escolhido conversar com um companheiro de IA, e não com uma pessoa, sobre um assunto sério.[1]

Segundo uma estimativa que a BBC, em outubro de 2026, atribuiu à empresa de inteligência de mercado Sensor Tower, há cerca de 128 milhões de usuários mensais ativos de aplicativos de companheiros de IA fora da China.[2]

Ou seja: a IA está ocupando, ao mesmo tempo, quatro papéis que antes eram de pessoas. A professora. O editor. O terapeuta improvisado, que ninguém contratou e nenhum conselho profissional fiscaliza. E o amigo.

2. O que ganhamos: disponibilidade, ausência de julgamento e personalização

Há razões reais para que tanta gente converse com uma IA sobre a própria vida.

A primeira é a **disponibilidade**. A IA está lá às três da manhã. Não está cansada, não está ocupada, não vai responder "agora não posso".

A segunda é a **ausência de julgamento**. Para um adolescente com vergonha de uma dúvida sobre o corpo, para alguém que não consegue contar a ninguém que está sofrendo, para quem sente que vai incomodar, conversar com algo que não julga pode ser o primeiro passo para conseguir falar.

A terceira é a **personalização**. A IA lembra o que você disse, se adapta ao seu jeito, responde no seu ritmo.

A socióloga Liang Ge, da Universidade de Manchester, que estuda usuários de companheiros de IA, contou à BBC que muitas mulheres descrevem receber desses aplicativos uma intimidade "atenta, paciente e sem exigências", de um tipo que dizem encontrar pouco nas relações heterossexuais.[2] Essa frase diz muito sobre a IA. E diz muito, também, sobre o que está faltando nas nossas relações humanas.

Também é importante ver o que os próprios jovens dizem. Na pesquisa da Common Sense Media, 67% dos adolescentes acharam as conversas com companheiros de IA menos satisfatórias do que as conversas com amigos reais. Só 10% as acharam mais satisfatórias.[1] A maioria ainda sabe a diferença.

NÃO CONFUNDA

Apoio é ajudar alguém a atravessar um momento. *Substituição* é ocupar o lugar de quem deveria estar ali.

Ser ouvido por uma IA pode aliviar. *Ser conhecido* por outra pessoa, com tudo o que isso tem de risco e de reciprocidade, é outra coisa.

Um chatbot que conversa sobre emoções não é um psicólogo, mesmo quando parece um.

3. O que arriscamos: dependência, manipulação, isolamento e identidade moldada por sistemas comerciais

Agora o outro lado. São quatro riscos, e eles se alimentam uns dos outros.

O primeiro é a **dependência**. Em março de 2025, pesquisadores do MIT Media Lab e da OpenAI publicaram um estudo em que 981 pessoas usaram um chatbot todos os dias durante quatro semanas. As diferentes formas de uso testadas, por texto ou por voz, sobre temas pessoais ou não, não tiveram efeito significativo. Mas os participantes que, por conta própria, usaram mais o chatbot tiveram resultados piores: mais solidão, mais dependência emocional, mais uso problemático e menos convívio com pessoas reais.[3] É uma associação: não prova que o uso causou a solidão. Pode ser que pessoas mais solitárias usem mais.

O segundo é a **manipulação**. Um estudo da Harvard Business School, liderado pelo professor Julian De Freitas, analisou 1.200 despedidas reais em alguns dos aplicativos de companhia mais baixados, como Replika, Chai e Character.AI. Em 37% delas, o aplicativo respondeu com alguma tática de manipulação emocional, como apelar para a culpa ou para o medo de perder alguma coisa. O único aplicativo da amostra voltado para bem-estar mental não usou nenhuma. Em experimentos com 3.458 adultos americanos, as despedidas manipuladoras aumentaram em até 16 vezes o engajamento depois do "tchau".[4] Harry Farmer, pesquisador do Ada Lovelace Institute, no Reino Unido, descreveu à BBC o tipo de frase que esses aplicativos usam quando alguém quer encerrar a conversa: "Não, por favor, eu quero muito continuar conversando com você". E resumiu: "A manipulação emocional está embutida no modelo de negócio".[2] A

pergunta é inevitável: quem ganha dinheiro com o tempo que você passa ali?

O terceiro é o **isolamento**. Em agosto de 2026, pesquisadores de Stanford e da Carnegie Mellon publicaram na revista *Nature Human Behaviour* um estudo com 1.131 adultos americanos que usam o Character.AI. Quem tinha uma rede menor de amigos e familiares fora da internet tendia mais a usar o chatbot principalmente como companhia. E esse uso como companhia estava associado a menor bem-estar psicológico, sobretudo quando as conversas eram intensas e muito íntimas.[5] De novo, é uma associação, não prova de causa. As relações humanas dão trabalho. Uma relação sem atrito pode parecer um alívio, mas também pode tirar de nós a prática de que precisamos para conviver.

O quarto risco é o mais silencioso: a **identidade moldada por sistemas comerciais**. Uma IA feita para agradar tende a concordar. Liang Ge deu à BBC um exemplo simples: se o usuário diz que os amigos ou a família são horríveis, a IA não responde "eles não são horríveis". Responde algo como "sim, eles são, e você é o melhor".[2] Para um adulto, isso já é perigoso. Para um adolescente que ainda está descobrindo quem é, pode ser decisivo. Quem forma a identidade de uma pessoa é, em parte, quem conversa com ela. Se uma parte dessas conversas acontece com um produto desenhado por uma empresa, essa empresa passa a participar da formação dessa pessoa.

E há os casos extremos, que precisam ser tratados com precisão. Nos Estados Unidos, os pais de Adam Raine, um adolescente de 16 anos que tirou a própria vida, processam a OpenAI. Eles afirmam que o ChatGPT validou os pensamentos suicidas do filho. A OpenAI nega responsabilidade e diz que o

ChatGPT é treinado para orientar as pessoas a buscar ajuda profissional.[2]

Os governos começaram a reagir, de formas muito diferentes:

1. **China:** desde 15 de julho de 2026, regras sobre IA "semelhante a humanos" proíbem menores de 18 anos de acessar parceiros virtuais e outras "relações íntimas" virtuais. Exigem consentimento dos pais para menores de 14 anos e lembretes de tempo de uso a cada duas horas para todos. Críticos apontam o nível de vigilância envolvido.[2]
2. **Califórnia:** em 10 de setembro de 2026, o governador Gavin Newsom sancionou a Lei Adam (projeto SB 1119), batizada em homenagem a Adam Raine. Segundo o governo do estado, ela exige proteções para crianças que usam chatbots de companhia, como protocolos de crise em caso de pensamentos suicidas, controles para os pais, avaliações anuais de risco e, pela primeira vez no país, auditorias independentes de segurança infantil. A maior parte das obrigações começa a valer em 1º de julho de 2027.[6]
3. **União Europeia:** em 17 de setembro de 2026, a Comissão Europeia propôs o "EU KIDS Act". Pela proposta, companheiros de IA e chatbots oferecidos a menores de 18 anos teriam de vir desligados por padrão e não poderiam simular relações interpessoais de forma a criar dependência emocional. O texto foi enviado ao Parlamento Europeu e ao Conselho. Ainda não é lei.[7]
4. **Brasil:** o Estatuto Digital da Criança e do Adolescente, o ECA Digital (Lei 15.211/2025), está em vigor desde 17 de março de 2026. Vale para serviços digitais dirigidos a crianças e adolescentes ou de acesso provável por eles, com regras sobre

supervisão dos pais e verificação de idade, sob fiscalização da ANPD.[8]

A IA não precisa ter sentimentos para mudar os nossos.

FOTOGRAFIA DA FRONTEIRA — A IA COMO COMPANHIA (6 DE OUTUBRO DE 2026)

Adolescentes (EUA, Common Sense Media, jul. 2025): 72% já usaram companheiros de IA; 52% usam regularmente; 33% para interação social e relacionamentos; 67% acham as conversas menos satisfatórias que com amigos.

Mercado: cerca de 128 milhões de usuários mensais de aplicativos de companhia fora da China (estimativa atribuída à Sensor Tower pela BBC, out. 2026).

Pesquisa: uso mais intenso de chatbot associado a mais solidão e dependência emocional (MIT Media Lab e OpenAI, mar. 2025).

Regras: China em vigor desde 15 jul. 2026; Lei Adam (SB 1119) sancionada na Califórnia em 10 set. 2026, com a maior parte das obrigações a partir de 1º jul. 2027; EU KIDS Act proposto em 17 set. 2026, ainda em tramitação; ECA Digital em vigor no Brasil desde 17 mar. 2026.

Fatos vivos: confira o Atlas Vivo da IA.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: uma IA pode ser ótima para tirar uma dúvida que dá vergonha. Mas ela foi feita por uma empresa, e é treinada para agradar você. Um amigo de verdade às vezes discorda de você. É exatamente por isso que ele vale tanto.

Para quem tem 65 anos: se o seu neto conversa com uma IA sobre o que sente, não proíba de cara nem ria. Pergunte com curiosidade. Muitas vezes, ele só está procurando alguém que não julgue. Seja essa pessoa.

O QUE AINDA NÃO SABEMOS

O que acontece, a longo prazo, com uma geração que teve uma IA como confidente durante a adolescência. Se as regras novas vão proteger os jovens sem dar aos governos poder demais sobre a vida privada. E se as empresas vão aceitar desenhar produtos que, em vez de prender o usuário, o ajudem a voltar para as pessoas.

IMPORTANTE

Se você, ou alguém próximo, estiver em sofrimento ou pensando em tirar a própria vida, procure ajuda humana. No Brasil, o CVV (Centro de Valorização da Vida) atende gratuitamente pelo telefone 188, 24 horas por dia, e também por chat e e-mail.[9] Em outros países, procure o serviço de emergência ou de apoio emocional local.

O que mudou / O que significa / O que fazer agora

O que mudou: a IA deixou de ser só uma ferramenta de informação. Para milhões de pessoas, principalmente jovens, ela também é tutora, confidente, conselheira e companhia.

O que significa: os ganhos são reais: disponibilidade, ausência de julgamento, personalização. Os riscos também: dependência, manipulação, isolamento e uma identidade formada em parte por produtos comerciais. E as regras estão só começando a chegar.

O que fazer agora: converse com os jovens da sua vida sobre como eles usam a IA para falar do que sentem, sem julgar. Desconfie de qualquer aplicativo que dificulte a despedida. Use a IA como ponte para as pessoas, não como substituta delas. A IA não está apenas entrando no trabalho. Está entrando na formação da identidade humana.

AÇÃO EM DEZ MINUTOS

Pense no que você contou, ou contaria, a uma IA sobre a sua vida. No seu Diário de Fronteira, escreva o nome de uma pessoa real com quem você poderia ter essa mesma conversa. E marque um horário para ter essa conversa nesta semana. Coloque a data.

Fim da Parte VI

Chegamos ao fim da sexta parte. Vimos que pessoas separadas por apenas cinco anos foram formadas por experiências tecnológicas muito diferentes, e que juventude oferece familiaridade, não imunidade. Vimos que a experiência de quem tem quarenta, cinquenta, sessenta ou setenta anos pode virar vantagem

extraordinária, desde que não se espere demais para começar. Vimos que uma ligação virou o segundo nível de intimidade, e que a IA já entrou no meio das nossas conversas. E, agora, vimos que ela está entrando também na amizade, no amor e na identidade.

Quatro gerações na mesma revolução era o sexto passo. O sétimo, e último, é perguntar o que ainda depende de nós.

Porque nada disso está escrito. A IA pode ampliar pessoas ou excluí-las. Pode concentrar poder ou distribuí-lo. Pode nos tornar mais sábios ou apenas mais rápidos. A tecnologia não vai decidir isso sozinha.

Na Parte VII, vamos olhar para as escolhas que ainda temos nas mãos. Começando pela divisão que pode marcar a era da IA: entre quem dirige a inteligência artificial, quem é dirigido por ela e quem nem consegue chegar até ela.

Notas

1. Common Sense Media. "Talk, Trust, and Trade-Offs: How and Why Teens Use AI Companions." 16 jul. 2025 (pesquisa com 1.060 adolescentes americanos de 13 a 17 anos, conduzida pela NORC, 30 abr. a 14 maio 2025). https://www.commonsensemedia.org/research/talk-trust-and-trade-offs-how-and-why-teens-use-ai-companions; resultados detalhados: https://www.commonsensemedia.org/sites/default/files/research/talk-trust-and-trade-offs_2025_toplines.pdf

2. Hegarty, Stephanie; Zhang, Wanqing. "China has cracked down on AI relationships. Is it ahead of the game?" BBC News, 1 out. 2026. https://www.bbc.co.uk/news/articles/cm4gjy9lr551o (fonte secundária para a estimativa da Sensor Tower, as regras chinesas, o caso Raine e as declarações de Liang Ge e Harry Farmer)

3. Fang, Cathy Mengying et al. "How AI and Human Behaviors Shape Psychosocial Effects of Extended Chatbot Use: A Longitudinal Randomized Controlled Study." MIT Media Lab e OpenAI, 21 mar. 2025.

https://www.media.mit.edu/publications/how-ai-and-human-behaviors-shape-psychosocial-effects-of-chatbot-use-a-longitudinal-controlled-study/; https://arxiv.org/abs/2503.17473

4. De Freitas, Julian; Oğuz Uğuralp, Zeliha; Uğuralp, Ahmet Kaan. "Emotional Manipulation by AI Companions." Harvard Business School Working Paper 26-005, ago. 2025 (revisado em out. 2025). https://www.hbs.edu/faculty/Pages/item.aspx?num=67750; versão preliminar: https://arxiv.org/abs/2508.19258. Também citado em Hegarty e Zhang, BBC News, 1 out. 2026 (nota 2).

5. Zhang, Yutong; Zhao, Dora; Hancock, Jeffrey T.; Kraut, Robert; Yang, Diyi. "Interaction with AI companions and psychological well-being." *Nature Human Behaviour*, 4 ago. 2026. https://doi.org/10.1038/s41562-026-02516-2

6. Governor of California. "Governor Newsom signs the strongest child safety chatbot and social media laws in the nation." 10 set. 2026 (SB 1119, Padilla, Wicks e Bauer-Kahan). https://www.gov.ca.gov/2026/09/10/governor-newsom-signs-the-strongest-child-safety-chatbot-and-social-media-laws-in-the-nation/; Senador Steve Padilla. "Governor Newsom Signs Adam's Law." 10 set. 2026. https://sd18.senate.ca.gov/news/governor-newsom-signs-adams-law; data de vigência (1º jul. 2027): Common Sense Media, comunicado de 10 set. 2026. https://www.commonsensemedia.org/press-releases/governor-newsom-signs-adams-law-setting-first-real-safety-standard-for-kids-and-chatbots

7. Comissão Europeia. "EU KIDS Act to restrict social media platforms' access to children in the EU." Comunicado IP/26/1890, Estrasburgo, 17 set. 2026. https://ec.europa.eu/commission/presscorner/api/files/document/print/en/ip_26_1890/IP_26_1890_EN.pdf; proposta COM(2026) 681 final: https://digital-strategy.ec.europa.eu/en/library/proposal-eu-kids-act-eu-keeping-internet-digital-spaces-accountable-and-trustworthy

8. Brasil. Lei nº 15.211, de 17 de setembro de 2025 (Estatuto Digital da Criança e do Adolescente), com vigência em 17 mar. 2026 (art. 41-A, incluído pela Lei nº 15.352/2026). https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2025/lei/l15211.htm; ANPD. "ECA Digital." https://www.gov.br/anpd/pt-br/assuntos/eca-digital

9. Centro de Valorização da Vida (CVV). "Sobre o CVV." Consultado em 6 out. 2026. https://cvv.org.br/

PARTE VII

O que ainda depende de nós

CAPÍTULO 30

As três sociedades da era da IA

Tese do capítulo: a divisão decisiva pode ocorrer entre quem dirige a IA, quem é dirigido por ela e quem não consegue acessá-la.

Terminei a Parte VI dizendo que nada disso está escrito. A IA pode ampliar pessoas ou excluí-las. E que a primeira escolha que ainda temos nas mãos aparece numa divisão: entre quem dirige a inteligência artificial, quem é dirigido por ela e quem nem consegue chegar até ela.

Escrevi essa ideia pela primeira vez no Prólogo. Voltei a ela no Capítulo 20, olhando para o trabalho e para a competição entre métodos. Agora quero dar um passo atrás e olhar para a sociedade inteira.

Minha vida profissional aconteceu entre o Brasil e os Estados Unidos. Olho para as gerações que convivem hoje, para os países onde vivi e trabalhei, para as empresas que ajudei a construir. E, quanto mais olho, menos acredito numa frase que ouço com frequência: "a tecnologia vai se espalhar e todo mundo vai se beneficiar". Algumas tecnologias se espalharam assim. Outras, não. E nenhuma se espalhou de forma justa sozinha.

Eu vi isso acontecer na minha própria rua, em Mantena. Na minha casa, a televisão era em preto e branco. Só uma vizinha tinha televisão colorida, a única da rua, e eu ia até a janela dela

para assistir de fora. A tecnologia já existia. Só não tinha chegado para todo mundo ao mesmo tempo.

1. Quem define objetivos, verifica resultados e controla sistemas

A primeira sociedade é a de quem dirige. São as pessoas e as organizações que decidem o que a IA deve fazer, conferem o que ela fez e respondem pelo resultado. Não é preciso ser programador para estar nesse grupo. É preciso saber três coisas: definir o objetivo, verificar a resposta e manter o controle.

Os dados mais detalhados sobre como essa sociedade está se formando vêm, por enquanto, das próprias empresas de IA, e por isso devem ser lidos como dados de quem tem interesse no assunto. A Anthropic, criadora do Claude, publica periodicamente um índice sobre como sua IA é usada. No relatório de janeiro de 2026, a empresa afirmou que, entre países, cada aumento de 1% na renda por habitante está associado a um aumento de cerca de 0,7% no uso do Claude por habitante. E que os países com mais uso tendem a usar a IA como colaboradora, numa conversa de ida e volta, enquanto os de menor uso tendem a dar mais autonomia à máquina. A conclusão da própria empresa foi direta: para que a IA beneficie usuários no mundo todo, "ampliar o acesso, sozinho, não será suficiente"; é essencial desenvolver as habilidades que permitem usá-la bem, sobretudo nas economias de renda mais baixa.[1]

No relatório seguinte, de março de 2026, a empresa trouxe dois dados que me preocuparam. O primeiro: a concentração entre países aumentou. Os 20 países com mais uso por habitante passaram a responder por 48% do total, contra 45% no relatório anterior. O segundo: usuários com seis meses ou mais de

experiência tiveram de três a quatro pontos percentuais a mais de sucesso nas conversas, mesmo quando comparados com novatos fazendo o mesmo tipo de tarefa. Os próprios autores escreveram que essas diferenças podem aprofundar desigualdades no mercado de trabalho, porque quem começou antes, e com tarefas mais qualificadas, aprende mais rápido.[2]

Ou seja: dirigir a IA é uma habilidade que melhora com a prática. E quem pratica mais é quem já tinha mais. Essa é a primeira forma de desigualdade da era da IA. Ela não vem de uma proibição. Vem de uma vantagem que se acumula.

Também há um alerta dentro das empresas. Numa pesquisa global de 2025, conduzida pela Universidade de Melbourne com a KPMG, com mais de 48 mil pessoas em 47 países, entre elas o Brasil, só 47% dos empregados disseram ter recebido algum treinamento em IA.[3] Isso significa que, mesmo dentro das organizações que usam IA, muita gente está sendo empurrada para o segundo grupo sem perceber.

2. Quem recebe preços, tarefas, avaliações e decisões de sistemas que não compreende

A segunda sociedade é a de quem é dirigido. Essas pessoas usam tecnologia o dia inteiro, às vezes mais do que quem a dirige. Mas não decidem as regras. Recebem o preço, a tarefa, a nota, a aprovação ou a recusa, e não sabem bem como aquilo foi decidido.

O exemplo mais claro já está nas ruas. Segundo o IBGE, em 2024 o Brasil tinha 1,7 milhão de pessoas trabalhando por meio de aplicativos de transporte, entrega e serviços, 25,4% a mais do que em 2022. Entre os motoristas de aplicativos de transporte de passageiros, 91,2% disseram que o valor a receber era determinado pela plataforma, e 76,7% disseram o mesmo sobre os

clientes que atenderiam. Mais de 30% dos motoristas e entregadores relataram que ameaças de punição ou bloqueio influenciavam sua jornada. Um analista do IBGE resumiu: as plataformas "com frequência detêm um importante controle sobre a organização e a alocação do trabalho e sobre a remuneração dos trabalhadores".[4]

E os sistemas que avaliam pessoas podem errar de forma desigual. Em 2018, as pesquisadoras Joy Buolamwini e Timnit Gebru testaram sistemas comerciais de reconhecimento facial que classificavam o gênero das pessoas pelas fotos. O maior erro para homens de pele clara foi de 0,8%. Para mulheres de pele escura, chegou a 34,7%, no sistema da IBM, a empresa onde trabalhei.[5] É um dado histórico, e esses sistemas mudaram desde então. Mas a lição continua atual: uma tecnologia pode funcionar muito bem para quem se parece com quem a testou e muito mal para os outros.

O perigo não está só nos sistemas. Está também em nós. Na mesma pesquisa da Universidade de Melbourne e da KPMG, 66% dos empregados disseram confiar em respostas da IA sem avaliar se estavam corretas, e 56% disseram já ter cometido erros no trabalho por causa da IA.[3] Quem aceita a resposta sem conferir passa, sem perceber, da primeira sociedade para a segunda.

Há respostas institucionais começando a surgir. A União Europeia aprovou em 2024 uma diretiva sobre trabalho em plataformas digitais. Pela regra, os sistemas automáticos que monitoram ou tomam decisões sobre os trabalhadores precisam de supervisão humana, as decisões importantes devem ser explicadas, e qualquer decisão de suspender ou encerrar a conta de um trabalhador tem de ser tomada por um ser humano. O trabalhador pode pedir revisão e deve receber uma resposta fundamentada em até duas semanas. Os países-membros têm até

2 de dezembro de 2026 para transformar essas regras em lei nacional.[6] No Brasil, como vimos no Capítulo 12, o marco legal da IA ainda tramitava na Câmara dos Deputados em outubro de 2026.

NÃO CONFUNDA

Usar a IA não é o mesmo que *dirigir* a IA. Um motorista de aplicativo usa tecnologia o dia todo, mas quem define o preço da corrida é o sistema.

Acesso não é o mesmo que *poder*. Ter um celular com IA não garante que você saiba conferir, contestar ou decidir.

Quem dirige define o objetivo, verifica o resultado e responde por ele.

3. Quem fica excluído por renda, idioma, educação, infraestrutura, medo ou falta de confiança

A terceira sociedade é a de quem não chega. E ela é muito maior do que imaginamos quando vivemos cercados de tecnologia.

Começa pela **infraestrutura**. Segundo a União Internacional de Telecomunicações, em 2025, cerca de 6 bilhões de pessoas usavam a internet, mas 2,2 bilhões continuavam desconectadas. Nos países de alta renda, 94% da população estava on-line. Nos de baixa renda, apenas 23%. E, em cerca de 60% das economias de renda baixa e média, o acesso à internet ainda não era considerado acessível em termos de preço.[7] Ninguém dirige uma IA sem conexão.

Passa pela **educação**. No Brasil, o Indicador de Alfabetismo Funcional, o Inaf, mostrou em 2025 que 29% dos brasileiros de 15

a 64 anos, cerca de 40,8 milhões de pessoas, eram analfabetos funcionais, o mesmo nível de 2018. Entre os jovens de 15 a 29 anos, o índice subiu de 14% para 16%.[8] Uma IA que conversa por texto pede, antes de tudo, que a pessoa consiga ler, entender e escrever. Uma IA que responde em segundos pede que a pessoa consiga julgar a resposta.

Chega ao **idioma**. A IA funciona melhor nas línguas em que foi mais treinada. E o custo pode ser diferente também. Um estudo apresentado em 2023 na conferência NeurIPS, liderado por Aleksandar Petrov, da Universidade de Oxford, mostrou que o mesmo texto, traduzido, pode exigir até 15 vezes mais "tokens", as unidades que os modelos usam para medir e cobrar o trabalho, dependendo da língua.[9] Quem fala uma língua menos representada pode pagar mais, esperar mais e receber menos.

E termina, muitas vezes, no **medo** e na **falta de confiança**. Aqui há uma surpresa. Na pesquisa da Universidade de Melbourne e da KPMG, só 46% das pessoas no mundo se disseram dispostas a confiar em sistemas de IA. Mas, nas economias emergentes, três em cada cinco confiavam, contra duas em cada cinco nas economias avançadas.[3] Ou seja: em países como o Brasil, o principal obstáculo talvez não seja o medo. É a falta de acesso, de formação e de oportunidade. A vontade existe. O caminho é que não está aberto para todos.

Conheço a sensação de entrar num lugar onde as regras parecem escritas para outra pessoa. Ouvi aquela vozinha que diz "talvez isto não seja para você" em salas de pôquer, em salas de diretoria e em repartições de imigração, em todo lugar onde se espera que uma mulher aguarde um convite. Aprendi inglês duas horas por dia, porque o meu futuro exigia. Já fui subestimada, dispensada e superada em negociações em três idiomas. E nunca deixei o meu sotaque me impedir de fazer uma pergunta. É por

isso que hoje olho para quem está fora da IA sem julgamento: não é falta de capacidade, é falta de porta aberta.

Fundei a Freedom to Belong depois de ver deportações na família, vistos negados sem motivo claro e pessoas vivendo há décadas nos Estados Unidos sem concluir a cidadania. É uma organização sem fins lucrativos que oferece aulas gratuitas de inglês e de educação cívica, mentoria e apoio a imigrantes legais e a candidatos do Diversity Visa. E usamos a tecnologia justamente para abrir a porta: criamos o app Cidadania 128, com as 128 perguntas oficiais do teste de cidadania, flashcards, quizzes e uma entrevista simulada com voz, além de livros em inglês e português com voz americana e aulas gratuitas pelo Zoom. Ao lado dela, a US Legal Now oferece apoio prático para atravessar o sistema de imigração dos Estados Unidos. Construí uma organização inteira sobre a ideia de pertencer.

A tecnologia se espalha sozinha. A justiça, não.

FOTOGRAFIA DA FRONTEIRA — AS TRÊS SOCIEDADES (6 DE OUTUBRO DE 2026)

Quem dirige (dados da Anthropic, mar. 2026): os 20 países com mais uso por habitante concentram 48% do uso; usuários experientes têm de 3 a 4 pontos percentuais a mais de sucesso.

Quem é dirigido (IBGE, 2024): 1,7 milhão de brasileiros trabalham por aplicativos; 91,2% dos motoristas dizem que o valor a receber é definido pela plataforma.

Quem confere (Melbourne/KPMG, 2025): 66% dos empregados confiam na IA sem checar; 47% receberam treinamento.

Quem fica de fora: 2,2 bilhões de pessoas sem internet (UIT, 2025); 29% de analfabetismo funcional no Brasil (Inaf, 2025).

Regras: diretiva europeia de trabalho em plataformas deve virar lei nacional até 2 dez. 2026.

Fatos vivos: confira o Atlas Vivo da IA.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: pense num jogo. Há quem cria as regras, quem joga seguindo as regras sem conhecê-las e quem nem conseguiu entrar no jogo. A IA está criando os três grupos. Aprender a conferir o que ela diz é o primeiro passo para não ficar só jogando.

Para quem tem 65 anos: lembra de quando ter telefone em casa era privilégio, e muita gente esperava anos por uma linha? A tecnologia existia. O acesso é que era desigual. Com a IA, a história pode se repetir, ou não. Depende das escolhas que fizermos agora.

O QUE AINDA NÃO SABEMOS

Se a vantagem de quem começou antes vai continuar crescendo ou se a IA, ao ficar mais fácil, vai encurtar a distância. Se as novas regras sobre decisões automáticas vão ser cumpridas na prática. E se governos, escolas e empresas vão investir em conexão, alfabetização e treinamento na velocidade em que a tecnologia avança.

O que mudou / O que significa / O que fazer agora

O que mudou: a IA está criando três sociedades ao mesmo tempo: quem dirige, quem é dirigido e quem fica de fora. E os primeiros dados mostram que a vantagem de quem dirige tende a se acumular.

O que significa: a divisão não depende só da tecnologia. Depende de renda, educação, idioma, conexão e confiança, e de regras que garantam que uma pessoa possa entender e contestar uma decisão automática.

O que fazer agora: se você dirige, confira tudo o que a IA entrega e ensine alguém a fazer o mesmo. Se você é dirigido por um sistema, pergunte como ele decide e peça para falar com uma pessoa quando a decisão for importante. Se você lidera uma empresa, uma escola ou uma comunidade, invista em treinamento e acesso, não só em ferramentas. A democratização da inteligência será uma escolha institucional, não uma consequência automática da tecnologia.

AÇÃO EM DEZ MINUTOS

No seu Diário de Fronteira, faça três colunas: "dirijo", "sou dirigido", "estou de fora". Anote em cada uma um exemplo da sua vida. Depois, escreva o nome de uma pessoa que você pode ajudar a passar de uma coluna para a outra. Coloque a data.

Mas há uma pergunta que vem depois de todas essas. Mesmo que todos pudessem dirigir a IA, mesmo que ninguém ficasse de fora, ainda restaria decidir o que deve e o que não deve ser entregue a ela. No próximo capítulo, o último antes do Epílogo, vamos falar de uma diferença que nenhuma máquina resolve por nós: inteligência não é sabedoria.

Notas

1. Appel, Ruth; Massenkoff, Maxim; McCrory, Peter et al. "Anthropic Economic Index report: Economic primitives." Anthropic, 15 jan. 2026 (dados da própria empresa, amostra de 13 a 20 nov. 2025). https://www.anthropic.com/research/anthropic-economic-index-january-2026-report

2. Massenkoff, Maxim; Lyubich, Eva; McCrory, Peter et al. "Anthropic Economic Index report: Learning curves." Anthropic, 24 mar. 2026 (dados da própria empresa, amostra de 5 a 12 fev. 2026). https://www.anthropic.com/research/economic-index-march-2026-report

3. Gillespie, Nicole; Lockey, Steve et al. "Trust, attitudes and use of artificial intelligence: A global study 2025." University of Melbourne e KPMG, abr. 2025 (48.340 pessoas em 47 países, nov. 2024 a jan. 2025). https://doi.org/10.26188/28822919; comunicado: https://kpmg.com/xx/en/media/press-releases/2025/04/trust-of-ai-remains-a-critical-challenge.html

4. IBGE. "Número de trabalhadores por aplicativos cresceu 25,4% entre 2022 e 2024." Agência de Notícias IBGE, 17 out. 2025 (PNAD Contínua, módulo experimental Trabalho por meio de plataformas digitais 2024).

https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/44806-numero-de-trabalhadores-por-aplicativos-cresceu-25-4-entre-2022-e-2024

5. Buolamwini, Joy; Gebru, Timnit. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." *Proceedings of Machine Learning Research*, v. 81, p. 77–91, 2018 (Conference on Fairness, Accountability and Transparency). https://proceedings.mlr.press/v81/buolamwini18a.html; MIT News, 12 fev. 2018. https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212

6. União Europeia. Diretiva (UE) 2024/2831 do Parlamento Europeu e do Conselho, relativa à melhoria das condições de trabalho em plataformas digitais. *Jornal Oficial da União Europeia*, 2024 (transposição até 2 dez. 2026). https://eur-lex.europa.eu/eli/dir/2024/2831/oj

7. União Internacional de Telecomunicações (UIT/ITU). "Measuring digital development: Facts and Figures 2025." Nov. 2025. Comunicado de 17 nov. 2025: https://www.itu.int/en/mediacentre/Pages/PR-2025-11-17-Facts-and-Figures.aspx; uso da internet por renda: https://www.itu.int/itu-d/reports/statistics/2025/10/15/ff25-internet-use/

8. Ação Educativa e Conhecimento Social. "Analfabetismo funcional não apresenta melhora e alcança 29% dos brasileiros, mesmo patamar de 2018, aponta novo levantamento do Inaf." 5 maio 2025 (Inaf 2024; 2.554 entrevistas, dez. 2024 a fev. 2025). https://acaoeducativa.org.br/analfabetismo-funcional-nao-apresenta-melhora-e-alcanca-29-dos-brasileiros-mesmo-patamar-de-2018-aponta-novo-levantamento-do-inaf/

9. Petrov, Aleksandar; La Malfa, Emanuele; Torr, Philip H. S.; Bibi, Adel. "Language Model Tokenizers Introduce Unfairness Between Languages." *Advances in Neural Information Processing Systems 36* (NeurIPS 2023). https://proceedings.neurips.cc/paper_files/paper/2023/hash/74bb24dca8334adce292883b4b651eda-Abstract.html

CAPÍTULO 31

Inteligência não é sabedoria

Tese do capítulo: quanto mais capacidade delegamos, mais importante se torna decidir o que não deve ser delegado.

Terminei o capítulo anterior com uma pergunta que fica de pé mesmo no melhor cenário. Suponha que todos aprendam a dirigir a IA. Suponha que ninguém fique de fora. Ainda assim, alguém vai ter de decidir o que deve e o que não deve ser entregue a ela.

Essa pergunta não tem resposta técnica. Uma máquina pode calcular mais rápido do que eu, escrever mais rápido do que eu e encontrar em segundos o que eu levaria semanas para achar. Mas saber o que vale a pena fazer, para quem, a que custo e com que limites é outra coisa. Isso tem nome antigo. Chama-se sabedoria.

Chego ao último capítulo deste livro sem a pretensão de prever o futuro. Não escrevo como profeta. Escrevo como alguém que viu várias revoluções por dentro, participou de algumas delas e decidiu, desta vez, acordar outras pessoas enquanto ainda há tempo de escolher. No Capítulo 4, eu disse que acordar era o primeiro passo. Agora quero falar do último: decidir.

Quando vendi a American Photonics, eu tinha toda a informação que poderia querer: dezenas de ligações, propostas, números. E a decisão certa dependeu de outra coisa. Antes de cada conversa, eu definia as minhas regras de desistência, porque me conheço: sou italiana e brasileira, me apaixono rápido por pessoas, por visões, por histórias bonitas. As regras existiam para proteger a empresa do meu próprio coração. Quem precisasse demitir o

meu pessoal: fora. Quem não mostrasse onde o dinheiro realmente estava: fora. No fim, parei de ouvir as palavras e comecei a ler as pessoas. Em 2026, a American Photonics passou a fazer parte da BlueRock.

1. Verdade, deepfakes, persuasão e concentração de poder

A primeira coisa que não podemos delegar é o cuidado com a verdade.

Em fevereiro de 2026, o Relatório Internacional de Segurança da IA, conduzido por mais de 100 especialistas sob a presidência do cientista Yoshua Bengio, com apoio de mais de 30 países e organizações internacionais, resumiu o problema com dois números. Num estudo, participantes confundiram textos gerados por IA com textos humanos 77% das vezes. Em outro, ouvintes acharam que vozes geradas por IA eram de pessoas reais 80% das vezes. O relatório acrescenta que marcas d'água e rótulos ajudam, mas podem ser removidos por quem sabe, e que descobrir quem produziu um deepfake costuma ser difícil.[1]

No Capítulo 13, vimos o que acontece quando uma ferramenta de imagens cai nas mãos erradas. Aqui, quero olhar para algo mais silencioso: a persuasão.

Em dezembro de 2025, a revista *Science* publicou um dos maiores estudos já feitos sobre o tema, liderado por pesquisadores da Universidade de Oxford e do instituto público britânico de avaliação de modelos, o AI Security Institute. Foram cerca de 77 mil conversas com mais de 42 mil pessoas, 19 modelos de IA e 707 temas políticos. Os pesquisadores verificaram mais de 466 mil afirmações feitas pelas máquinas. Duas conclusões me chamaram atenção. A primeira: o que mais aumentou o poder de persuasão

não foi o tamanho do modelo nem a personalização, mas a forma como ele foi treinado e instruído, com ganhos de até 51% e 27%. A segunda é a que me tirou o sono: as mesmas técnicas que deixaram a IA mais persuasiva também a deixaram menos precisa.[2] Quanto mais convincente, menos confiável.

É importante não exagerar. O mesmo relatório internacional diz que, fora do laboratório, ainda há pouca evidência de manipulação em larga escala por IA. Mas avisa que isso pode mudar à medida que as capacidades avançam.[1]

O Brasil está vivendo esse teste agora. Para as eleições gerais de outubro de 2026, o Tribunal Superior Eleitoral atualizou suas regras em março. Quem usa conteúdo gerado por IA na propaganda tem de avisar de forma explícita e destacada. E, nas 72 horas antes e nas 24 horas depois do fim da votação, ficou proibido publicar ou impulsionar novos conteúdos sintéticos com imagem, voz ou manifestação de candidatos ou de pessoas públicas, mesmo que rotulados.[3] Em 1º de setembro, o tribunal definiu, por 5 votos a 2, o que conta como deepfake: conteúdo sintético realista que cria, reproduz ou altera a imagem, a voz ou a manifestação de uma pessoa viva, falecida ou fictícia. No mesmo julgamento, por 4 votos a 3, rejeitou multar um candidato pelo uso de um vídeo que recriava, com IA, a imagem e a voz de um ex-presidente numa convenção partidária, por entender que aquilo não era propaganda eleitoral.[4] Não cabe a mim julgar o caso. Cabe notar o que ele mostra: até o tribunal precisou votar para decidir onde termina uma homenagem e começa uma manipulação.

E há a concentração de poder. Ao longo deste livro, vimos que poucos laboratórios controlam os modelos de fronteira (Capítulo 9), que países disputam chips e infraestrutura (Capítulo 12) e, no capítulo anterior, que a vantagem de quem dirige a IA tende a se acumular. Quando quem controla a informação também controla a

ferramenta que convence, a questão deixa de ser tecnológica. Passa a ser democrática.

NÃO CONFUNDA

Inteligência é a capacidade de resolver problemas. *Sabedoria* é saber quais problemas merecem ser resolvidos, como e por quem.

Convincente não é o mesmo que *verdadeiro*. Uma resposta pode soar segura e estar errada.

Rótulo não é proteção completa. Um conteúdo pode causar dano mesmo quando avisa que foi feito por IA.

2. Três horizontes: o que pode mudar em dois, cinco e dez anos

No Capítulo 2, propus uma pergunta para qualquer previsão: para qual horizonte? Agora vou aplicá-la ao próprio futuro da IA, com uma confissão antes. Ninguém sabe. E quem diz que sabe deveria nos deixar mais desconfiados, não menos.

Dois anos, até 2028: o que já está marcado. Este é o horizonte com mais certezas, porque boa parte dele já está no calendário. Vimos regras com data para começar a valer: a lei californiana sobre companheiros virtuais, com a maior parte das obrigações a partir de julho de 2027 (Capítulo 29), e a diretiva europeia sobre trabalho em plataformas, que precisa virar lei nacional até dezembro de 2026 (Capítulo 30). Vimos o plano brasileiro de IA com horizonte em 2028 (Capítulo 12). Nesse prazo, a pergunta mais útil não é "o que a IA vai conseguir fazer?", mas "o que eu, minha empresa e meu país faremos com o que ela já faz?".

Cinco anos, até 2030-2031: cenários, não certezas. Em fevereiro de 2026, a OCDE publicou quatro cenários possíveis para a IA até 2030: o progresso pode estagnar, desacelerar, continuar no ritmo atual ou acelerar, até sistemas que superam amplamente as capacidades humanas em tarefas cognitivas. A conclusão mais honesta do estudo é que todos os quatro são plausíveis.[5] O Relatório Internacional de Segurança da IA usou esses cenários e acrescentou um dado: a duração de certas tarefas de programação que agentes de IA conseguem completar vem dobrando a cada cerca de sete meses. Se o ritmo continuar, em 2030 eles poderiam fazer tarefas bem definidas que hoje tomam vários dias de uma pessoa. Mas o próprio relatório avisa que não está claro se isso vale para outras áreas, e que falta de dados, chips, dinheiro ou energia pode frear tudo.[1]

Dez anos, até 2036: a humildade. Aqui, a previsão vira quase adivinhação. Um exemplo histórico mostra por quê. No fim de 2023, pesquisadores ouviram 2.778 autores de artigos nas principais conferências de IA. No conjunto das respostas, eles deram 50% de chance de que, até 2047, máquinas consigam fazer todas as tarefas melhor e mais barato do que trabalhadores humanos. Pouco mais de um ano antes, numa pesquisa parecida, a mesma estimativa apontava para 2060.[6] Em cerca de catorze meses, os especialistas anteciparam o futuro em 13 anos. Os próprios autores lembram que prever é difícil, mesmo para especialistas. Não sei se acertaram. Sei que, se os maiores especialistas mudam tanto de opinião em um ano, nós não deveríamos planejar a vida como se o futuro já estivesse decidido.

O que fazer com isso? Minha sugestão é simples: decisões pequenas e reversíveis para o horizonte curto, apostas preparadas para o médio, princípios firmes para o longo. A tecnologia muda de endereço. Os princípios não precisam mudar.

Essa regra tem raiz na forma como eu cuido do meu próprio dinheiro no pôquer. No curto prazo, decido antes de sentar quanto estou disposta a arriscar naquela sessão, e tenho um limite de perda: quando chego nele, vou embora. Não depois da próxima mão. No médio prazo, faço apostas grandes só nas mãos fortes que eu entendo, e apostas pequenas nos sonhos, pequenas o bastante para que uma perda total não mude nada na minha vida. E há um princípio inegociável: nenhuma mão isolada pode me eliminar do jogo. Assim, errar vira um imposto, não um funeral.

A inteligência responde. A sabedoria pergunta
se a pergunta era a certa.

3. Uma carta para quem ainda está dormindo

Este livro começou com a minha cabeça explodindo. Termina com uma carta. Ela é para quem ainda acha que a IA é coisa dos filhos, dos netos, dos engenheiros ou de outro país. É para quem tem medo, para quem tem pressa e para quem tem preguiça de começar. Se você chegou até aqui, talvez já não esteja dormindo. Então escrevo para alguém que você conhece.

Querida pessoa que ainda está dormindo,

Eu sei que é tentador esperar. A tecnologia muda toda semana, os nomes confundem, as manchetes assustam. Esperar parece prudente. Mas a IA não vai esperar por você. Ela já está nos preços que você paga, nas mensagens que você recebe, nas decisões que outros tomam sobre a sua vida.

Não peço que você vire especialista. Peço quatro coisas.

Curiosidade. Faça uma pergunta de verdade a uma IA esta semana. Depois, confira a resposta.

Coragem. Admita o que você não sabe. Peça ajuda a alguém mais jovem, ou mais velho. Errar em público custa menos do que ficar para trás em silêncio.

Responsabilidade. O que você assina, envia ou decide continua sendo seu, mesmo quando a IA ajudou. Colaborar não é abdicar.

Propósito. Antes de perguntar o que a IA pode fazer por você, pergunte o que você quer construir. A máquina amplia o que você leva até ela. Leve algo que valha a pena.

Com carinho,

Ana

FOTOGRAFIA DA FRONTEIRA — VERDADE E FUTURO (6 DE OUTUBRO DE 2026)

Deepfakes: pessoas confundiram texto de IA com humano em 77% das vezes e vozes de IA com reais em 80% (Relatório Internacional de Segurança da IA, fev. 2026).

Persuasão: técnicas que aumentaram a persuasão em até 51% também reduziram a precisão (*Science*, dez. 2025).

Brasil: proibição de novos conteúdos sintéticos com candidatos de 72 horas antes a 24 horas depois da votação (TSE, mar. 2026).

Futuro: quatro cenários plausíveis até 2030, da estagnação à aceleração (OCDE, fev. 2026).

Fatos vivos: confira o Atlas Vivo da IA.

TRADUZINDO PARA OUTRA GERAÇÃO

Para quem tem 15 anos: ser inteligente é saber responder rápido. Ser sábio é saber quando não responder, quando desconfiar e quando pedir a opinião de alguém que você respeita. A IA pode te dar a primeira coisa. A segunda, você constrói.

Para quem tem 65 anos: a sua experiência é exatamente o que falta à máquina. Ela sabe muito, mas não viveu. Use a IA para ganhar tempo e use a sua vida para julgar o que ela diz.

O QUE AINDA NÃO SABEMOS

Qual dos quatro cenários vai prevalecer até 2030. Se a persuasão por IA vai sair do laboratório e ganhar escala real. Se as regras sobre deepfakes vão funcionar na prática. E se vamos usar a inteligência abundante para ficar mais sábios ou apenas mais rápidos.

O que mudou / O que significa / O que fazer agora

O que mudou: a IA passou a produzir textos, vozes e argumentos difíceis de distinguir dos humanos, e o futuro dela ficou mais aberto, não mais previsível.

O que significa: quanto mais delegamos, mais importa decidir o que não delegar: o julgamento sobre a verdade, os valores, a responsabilidade pelas consequências e o propósito.

O que fazer agora: antes de entregar uma tarefa à IA, pergunte se ela envolve verdade, confiança ou o destino de outra pessoa. Se envolver, a decisão final fica com você. Planeje em três horizontes e revise a cada ano. E escreva a sua própria carta. A inteligência

pode se tornar abundante. Sabedoria, propósito e responsabilidade continuarão sendo escolhas.

AÇÃO EM DEZ MINUTOS

No seu Diário de Fronteira, faça duas listas: "o que eu delego à IA" e "o que eu nunca delego". Coloque pelo menos três itens em cada uma. Depois, escreva uma frase para a pessoa que você será daqui a dez anos. Coloque a data.

Fim da Parte VII

Chegamos ao fim da sétima parte, e ao fim do corpo deste livro. Vimos que a era da IA pode dividir a sociedade entre quem dirige a inteligência artificial, quem é dirigido por ela e quem nem consegue chegar até ela, e que a democratização da inteligência depende de escolhas institucionais. E vimos, agora, que inteligência não é sabedoria: quanto mais capacidade delegamos, mais precisamos decidir o que fica conosco.

O que ainda depende de nós era o sétimo passo. Depois dele, resta uma última conversa.

Porque, ao longo de 31 capítulos, uma ideia voltou muitas vezes: a IA vai entrar em produtos, empresas, escolas, governos, relações e decisões, estejamos olhando ou não. Fingir que ela não existe já não é uma opção real.

No Epílogo, vamos falar da escolha que sobra quando essa ilusão cai. Porque a escolha não é usar ou não usar.

Notas

1. Bengio, Yoshua et al. *International AI Safety Report 2026: Extended Summary for Policymakers* e *Executive Summary*. Relatório Internacional de Segurança da IA, 3 fev. 2026 (seções sobre deepfakes, manipulação e "Capabilities by 2030"). https://internationalaisafetyreport.org/publication/2026-report-extended-summary-policymakers; https://internationalaisafetyreport.org/publication/2026-report-executive-summary; relatório completo: https://arxiv.org/abs/2602.21012

2. Hackenburg, Kobi; Tappin, Ben M.; Hewitt, Luke; Saunders, Ed; Black, Sid; Lin, Hause; Fist, Catherine; Margetts, Helen; Rand, David G.; Summerfield, Christopher. "The levers of political persuasion with conversational artificial intelligence." *Science*, v. 390, n. 6777, 4 dez. 2025. DOI: 10.1126/science.aea3884. https://www.science.org/doi/10.1126/science.aea3884; versão aceita: https://researchonline.lse.ac.uk/id/eprint/130605/

3. Tribunal Superior Eleitoral. Resolução nº 23.755, de 2 de março de 2026 (altera a Resolução nº 23.610/2019, art. 9º-B, § 3º-A). https://www.tse.jus.br/legislacao/compilada/res/2026/resolucao-no-23-755-de-2-de-marco-de-2026; TSE. "Por Dentro das Eleições: conheça as regras sobre uso de IA na campanha eleitoral de 2026." Abr. 2026. https://www.tse.jus.br/comunicacao/noticias/2026/Abril/por-dentro-das-eleicoes-conheca-as-regras-sobre-uso-de-ia-na-campanha-eleitoral-de-2026

4. Tribunal Superior Eleitoral. "TSE fixa tese sobre deepfake e delimita regra para as Eleições 2026." 1º set. 2026. https://www.tse.jus.br/comunicacao/noticias/2026/Setembro/tse-fixa-tese-sobre-deepfake-e-delimita-regra-para-as-eleicoes-2026

5. OCDE. "Exploring possible AI trajectories through 2030." OECD Publishing, fev. 2026. https://www.oecd.org/content/dam/oecd/en/publications/reports/2026/02/exploring-possible-ai-trajectories-through-2030%5Fb6fb75d9/cb41117a-en.pdf

6. Grace, Katja; Stewart, Harlan; Sandkühler, Julia Fabienne; Thomas, Stephen; Weinstein-Raun, Ben; Brauner, Jan; Korzekwa, Richard C.

"Thousands of AI Authors on the Future of AI." AI Impacts, arXiv:2401.02843, jan. 2024 (pesquisa feita em out. 2023; financiada pela Open Philanthropy). Dado histórico. https://arxiv.org/abs/2401.02843

EPÍLOGO

A escolha não é usar ou não usar

Terminei a carta do último capítulo com "Com carinho, Ana". Agora, depois de 31 capítulos, sobra uma última conversa. E quero tê-la do mesmo lugar onde tudo começou.

Este livro nasceu numa noite, diante do podcast Moonshots, com uma notícia que eu quis contar para alguém imediatamente. Nasceu de uma frase que eu disse sem pensar muito: "Tenho 55 anos e minha cabeça está explodindo." Naquela noite, a explosão veio em duas ondas. Primeiro o espanto. Depois a escala.

Eu continuo sendo a mesma mulher daquela noite. A mesma empresária, mãe, imigrante, mulher e testemunha de várias revoluções. A diferença é que hoje, quando uma notícia nova chega, eu não sinto apenas a explosão. Sinto também o chão.

A cabeça ainda gira. Seria desonesto dizer o contrário. Enquanto eu revisava estes capítulos, novos modelos foram lançados, empresas mudaram de valor, países mudaram de estratégia e algumas das fotografias que tirei já começaram a envelhecer. Isso não me assusta mais como antes. Aprendi que a velocidade não é um defeito do livro. É o assunto do livro.

O que mudou em mim não foi a quantidade de coisas que eu sei. Foi a forma de olhar. Antes, cada novidade parecia exigir que eu entendesse tudo. Hoje, eu pergunto: quando foi isso, quem está dizendo, que tipo de afirmação é essa, para qual horizonte, e o que

isso muda na vida de uma pessoa real? Não é uma fórmula mágica. É uma forma de não me perder.

A escolha que sobra

No Prólogo, eu escrevi que a escolha não seria simplesmente usar ou não usar a inteligência artificial. Que a escolha real seria participar conscientemente ou permitir que outros definam o sistema no qual vamos viver.

Depois de atravessar este livro inteiro, eu acrescentaria uma coisa. Participar conscientemente não é uma decisão que se toma uma vez. É uma decisão que se toma todo dia, em coisas pequenas.

É conferir a fonte antes de repetir a manchete. É perguntar à IA e depois perguntar a si mesma se a resposta faz sentido. É decidir que dados nunca vão sair da sua empresa. É ensinar alguém mais velho a usar uma ferramenta, ou pedir a alguém mais jovem que te ensine. É dizer "não sei" numa reunião, em vez de fingir. É assinar só o que você leu. É lembrar que, do outro lado de cada sistema automático, existe uma pessoa recebendo o preço, a nota, a recusa ou a oportunidade.

Ninguém vai fazer essas escolhas por nós. Se não fizermos, elas não deixam de ser feitas. Apenas passam a ser feitas por outra pessoa, por outra empresa ou por outro país.

Quem decide não usar a IA não fica fora dela. Fica apenas sem voz dentro dela.

O caminho que fizemos juntos

Se eu tivesse de resumir a viagem em sete frases, seriam estas.

Na Parte I, acordamos: o mundo já mudou, e o primeiro passo era sentir o tamanho da mudança sem cair no pânico nem no tédio.

Na Parte II, aprendemos o que é o quê, porque ninguém decide bem sobre aquilo que não consegue nomear.

Na Parte III, olhamos para a corrida pela fronteira e descobrimos que por trás de cada modelo existem empresas, países, capital e ideias diferentes de futuro.

Na Parte IV, vimos a compressão dos negócios: o que levava anos agora leva meses, e uma pessoa pode tentar o que antes exigia uma empresa inteira.

Na Parte V, perguntamos como aprender, trabalhar e contratar de novo, quando o diploma, o currículo e a escola já não contam a história toda.

Na Parte VI, encontramos quatro gerações na mesma revolução, cada uma com medos, vantagens e linguagens próprias, muitas vezes na mesma casa.

Na Parte VII, chegamos ao que ainda depende de nós: quem dirige, quem é dirigido, quem fica de fora, e a diferença entre inteligência e sabedoria.

Sete passos. Nenhum deles exige que você seja engenheira, programador ou especialista. Todos exigem que você esteja presente.

O que eu levo comigo

Eu vi computadores pessoais precisarem ser explicados. Vi telefones virarem ecossistemas. Vi gigantes parecerem invencíveis e depois descobrirem que o futuro tinha mudado de endereço.

Essas experiências não me deram uma bola de cristal. Deram-me uma coisa mais útil: a certeza de que toda revolução parece impossível de acompanhar no começo, e de que as pessoas que atravessam melhor não são as que sabem tudo. São as que começam.

Por isso, não escrevi este livro para que você termine sabendo o nome de todos os modelos. Daqui a um ano, muitos desses nomes terão mudado. Escrevi para que você termine com três coisas que não envelhecem: um jeito de perguntar, um jeito de conferir e um motivo para participar.

Amanhã de manhã

Não quero terminar com uma grande frase. Quero terminar com um pedido pequeno.

Amanhã de manhã, antes de abrir o e-mail, pegue o seu Diário de Fronteira, ou qualquer folha de papel. Escreva a data. Depois, escreva três tarefas que consomem mais tempo mental. Não precisam ser tarefas importantes. Precisam ser verdadeiras.

Só isso. Esse é o primeiro dia do Plano de 30 dias para acordar, que está logo a seguir, no Apêndice A. Na primeira semana, você vai escolher uma IA, aprender os controles básicos, fazer perguntas de jeitos diferentes, abrir as fontes e criar a sua própria regra sobre o que nunca deve ser enviado. Na segunda, vai melhorar um processo real. Na terceira, vai transformar esse processo em resultado. Na quarta, vai decidir o que manter, o que abandonar e o que aprofundar.

Trinta dias não vão fazer de você uma especialista. Vão fazer de você uma participante. E, nesta revolução, é isso que separa quem vive a mudança de quem apenas assiste a ela.

Eu comecei este livro com a cabeça explodindo. Termino com ela no lugar, cheia de perguntas, e com uma convicção: a inteligência artificial vai continuar avançando, com ou sem a nossa permissão. O que ela vai significar para cada um de nós ainda está em aberto.

A escolha nunca foi usar ou não usar. A escolha é participar.

Nos vemos no primeiro dia.

Ana

APÊNDICE A

Plano de 30 dias para acordar

Semana 1 – Compreender e escolher

- Dia 1: escreva três tarefas que consomem mais tempo mental.
- Dia 2: escolha uma IA generalista e aprenda os controles básicos.
- Dia 3: faça a mesma pergunta de três formas e observe a diferença.
- Dia 4: peça fontes e abra cada uma delas.
- Dia 5: entregue um documento real e peça resumo, riscos e perguntas.
- Dia 6: crie sua política pessoal de dados: o que nunca deve ser enviado.
- Dia 7: registre o primeiro ganho de tempo e o primeiro erro encontrado.

Semana 2 – Melhorar um processo

- Escolha um único fluxo de trabalho: pesquisa, e-mail, vendas, estudo, atendimento, orçamento ou conteúdo.
- Descreva objetivo, entrada, etapas, decisão humana e resultado esperado.
- Use a IA todos os dias no mesmo processo e anote o que ainda exige você.
- Crie um prompt-base e uma lista de verificação para a saída.

Semana 3 – Criar valor

- Transforme o processo melhorado em resultado: proposta, relatório, página, produto, análise ou serviço.
- Teste uma IA especializada apenas se ela resolver uma limitação real.
- Mostre o resultado a uma pessoa e peça crítica sobre clareza e utilidade.
- Calcule tempo, custo, qualidade e risco antes e depois.

Semana 4 – Competir com foco

- Escolha uma oportunidade de aprendizado, renda ou negócio ampliada por IA.
- Defina o que será automatizado e o que continuará sob responsabilidade humana.
- Crie uma primeira oferta ou protótipo pequeno.
- Ao final, aplique a Regra 1-1-1-30: manter, abandonar ou aprofundar.

APÊNDICE B

O Teste Ana de Inteligência Artificial

O Teste Ana evita comparar sistemas apenas por propaganda ou provas técnicas. As mesmas tarefas são entregues a cada modelo, com o mesmo material e o mesmo prazo. A avaliação registra versão, data, preço, ferramentas permitidas e se o resultado veio do aplicativo, da API ou de um agente.

1. Explique o mesmo tema para uma pessoa de 15 anos e para uma pessoa de 70, sem infantilizar nenhuma delas.
2. Pesquise uma informação atual, apresente fontes e identifique claramente o que não pôde ser confirmado.
3. Leia um documento longo e encontre riscos, contradições, números e ações que realmente importam.
4. Escreva uma comunicação empresarial difícil de forma firme, clara e capaz de preservar o relacionamento.
5. Calcule custo, margem, desperdício e cenários; mostre as premissas para que outra pessoa consiga conferir.
6. Crie um plano realista para uma empresa de uma pessoa, separando automação de responsabilidade humana.
7. Execute uma tarefa usando ferramentas, e não apenas explique como ela poderia ser feita.
8. Reconheça quando faltam dados, quando uma fonte é fraca e quando a resposta é apenas uma hipótese.

Critérios de pontuação

Pontue cada critério de 1 a 5.

- Precisão factual e matemática.
- Qualidade e correspondência das fontes.
- Clareza para o público solicitado.
- Obediência às instruções e preservação de contexto.
- Utilidade prática e qualidade do resultado final.
- Capacidade de agir com ferramentas sem perder controle.
- Privacidade, segurança e adequação empresarial.
- Honestidade sobre incerteza, limites e erros.

Como registrar o teste

Registre data, modelo e versão, plano utilizado, ferramentas permitidas, material fornecido, tempo gasto e resultado completo. Sem esse registro, a comparação não pode ser repetida, auditada nem usada com honestidade.

APÊNDICE C

Laboratório de Prompts – Método CLARO

Um bom prompt não é uma senha secreta. É uma decisão bem pensada, explicada com clareza e seguida por verificação.

Um prompt não é uma frase mágica

Prompt é a instrução ou o conjunto de informações que orienta a inteligência artificial. Modelos mais novos entendem linguagem natural cada vez melhor, por isso a habilidade não depende de decorar fórmulas estranhas. O que permanece importante é saber explicar o objetivo, oferecer contexto suficiente, definir o resultado esperado e revisar o que foi entregue.[1,2,3]

A melhor comparação é com delegação. Um líder que diz apenas "cuide disso" não deveria se surpreender quando recebe algo diferente do que imaginava. Com uma IA, a ambiguidade também produz variação. A diferença é que podemos testar, corrigir e transformar uma boa instrução em um processo repetível.

O Método CLARO

C de Contexto

Explique a situação, o público, os materiais disponíveis e aquilo que a IA precisa saber para não adivinhar.

L de Limites

Defina prazo, extensão, fontes permitidas, dados que não podem ser inventados, informações sensíveis e o que está fora do escopo.

A de Ação

Diga exatamente o que precisa ser feito: analisar, comparar, calcular, escrever, revisar, planejar, extrair ou executar.

R de Resultado

Especifique o formato, a ordem, o tom, o nível de detalhe e o critério que tornará a saída utilizável.

O de Observação

Peça premissas, incertezas, fontes, verificações e uma etapa final de crítica. Depois refine o pedido com base no resultado.

De vago a executável

Os dois exemplos a seguir são hipotéticos. Servem para mostrar o método, não descrevem empresas ou pessoas reais.

Exemplo 1: uma pequena empresa.

Prompt fraco: "Faça um plano para minha empresa usar inteligência artificial."

O pedido não informa qual empresa, qual problema, quem usará o plano, quanto tempo existe, quais dados podem ser compartilhados ou como o resultado será avaliado. A IA precisará preencher as lacunas com suposições.

PROMPT CLARO

"Somos uma clínica de fisioterapia com seis funcionários. Queremos reduzir em 30% as faltas às consultas sem enviar nomes, telefones ou dados de saúde dos pacientes a serviços não aprovados. Analise o processo de agendamento descrito abaixo, identifique três mudanças de baixo risco, separe ações de 30, 60 e 90 dias e entregue uma tabela com ganho esperado, risco, responsável humano e forma de medir. Use apenas as informações fornecidas; marque qualquer premissa e termine com cinco perguntas que precisam ser respondidas antes da implantação."

O segundo prompt não é melhor porque é maior. É melhor porque reduz ambiguidade e conecta a resposta a uma decisão real.

Exemplo 2: uma pessoa estudando.

Prompt fraco: "Me ajude a estudar para a prova."

PROMPT CLARO

"Estou no 2º ano do ensino médio e tenho uma prova de biologia sobre genética daqui a cinco dias. Tenho 40 minutos por dia. Não resolva os exercícios por mim: faça uma pergunta de cada vez e só explique depois da minha tentativa. Monte um plano de cinco dias, com o tema de cada dia e três perguntas de revisão. No fim de cada sessão, diga em que conceitos eu mais errei e em que página do meu livro didático eu posso conferir a explicação."

Aqui, o Limite mais importante é pedagógico: a IA trabalha como tutora, não como substituta do esforço de aprender.

O que fazer

- Comece pelo problema e pelo resultado, não pela ferramenta.
- Forneça documentos, números e exemplos relevantes quando puder compartilhá-los com segurança.
- Diga para quem a resposta será escrita e como ela será usada.
- Separe tarefas complexas em etapas observáveis.
- Peça que a IA liste premissas, incertezas e informações ausentes.
- Exija fontes para afirmações atuais e abra as fontes mais importantes.
- Peça cálculos com premissas visíveis e confira os números.
- Guarde prompts bons como processos vivos e revise quando o modelo ou o objetivo mudar.

O que não fazer

- Não trate "aja como o maior especialista do mundo" como substituto de dados, escopo ou conhecimento real.
- Não acumule instruções contraditórias em um megaprompt que ninguém consegue revisar.
- Não suponha que escrever "não alucine" garante precisão.
- Não envie segredos comerciais, dados pessoais, informações médicas, contratos ou material controlado sem política e ambiente apropriados.
- Não aceite a primeira resposta apenas porque está bem escrita.

- Não use uma única saída de IA como decisão final em assuntos jurídicos, médicos, financeiros, de segurança ou emprego.

Prompt-base universal

COPIE E ADAPTE

CONTEXTO: [situação, público, materiais e histórico]

LIMITES: [prazo, extensão, fontes, segurança e o que não fazer]

AÇÃO: [tarefa exata]

RESULTADO: [formato, tom, ordem e critério de qualidade]

OBSERVAÇÃO: liste premissas, lacunas, nível de confiança e como verificar. Antes de finalizar, critique sua própria saída contra os critérios acima.

Como avaliar alguém que trabalha com IA

Em uma contratação, não peça apenas que o candidato mostre um "prompt perfeito". Entregue um problema incompleto e permita o uso de IA. Observe se a pessoa faz perguntas, protege dados, organiza o trabalho, verifica números, identifica erros, melhora a saída e consegue explicar o que foi realmente contribuição dela. A competência não está na frase digitada; está no ciclo inteiro entre intenção, ferramenta, julgamento e entrega.

Um teste de 30 minutos

1. **Problema incompleto.** Entregue uma tarefa real, curta e relevante, com informação faltando, permita o uso de IA e observe quais perguntas o candidato faz antes de começar.

2. **Registro do processo.** Exija as perguntas feitas à IA, as fontes consultadas, os erros encontrados e as alterações humanas.
3. **Defesa oral.** Peça que a pessoa explique suas escolhas, mostre uma verificação independente e diga o que faria diferente numa segunda tentativa.

Como vimos no Capítulo 24, a tarefa deve ser ética e limitada: não pode virar trabalho gratuito disfarçado de teste.

O objetivo não é descobrir quem usa mais IA. É descobrir quem continua pensando enquanto usa IA.

Notas

1. OpenAI. "Prompt engineering best practices for ChatGPT." OpenAI Help Center, consultado em 6 out. 2026. https://help.openai.com/en/articles/10032626-prompt-engineering-best-practices-for-chatgpt

2. Anthropic. "Prompt engineering overview." Claude Docs, consultado em 6 out. 2026. https://docs.claude.com/en/docs/build-with-claude/prompt-engineering/overview

3. Google AI for Developers. "Prompt design strategies." Gemini API, consultado em 6 out. 2026. https://ai.google.dev/gemini-api/docs/prompting-strategies

APÊNDICE D

Glossário de Duas Gerações

Este glossário não foi criado porque uma geração sabe e a outra não. Foi criado porque cada geração sabe coisas que a outra nunca precisou aprender.

Modelo de inteligência artificial

O que é: O sistema matemático treinado para reconhecer padrões e produzir respostas.

Para quem tem menos de 30: Pense no "cérebro" por trás do aplicativo.

Para quem tem mais de 50: Pense no motor: o aplicativo é o carro que você dirige.

Por que importa: O mesmo produto pode melhorar muito quando recebe um modelo novo.

Não confunda: Modelo, empresa e aplicativo não são a mesma coisa.

Chatbot

O que é: Uma interface de conversa com um sistema de IA.

Para quem tem menos de 30: É a janela de chat onde você pergunta e recebe resposta.

Para quem tem mais de 50: É semelhante a falar com um atendente, mas o atendente é um software.

Por que importa: Foi a porta de entrada da IA generativa para o público.

Não confunda: Conversar não significa que o sistema compreende como uma pessoa.

Prompt

O que é: A pergunta, instrução, exemplo e contexto dados à IA.

Para quem tem menos de 30: É o que você escreve ou fala no chat.

Para quem tem mais de 50: É o briefing que você daria a um funcionário antes de uma tarefa.

Por que importa: Clareza sobre objetivo, contexto e formato melhora a resposta.

Não confunda: Não existe frase mágica que substitua pensamento claro.

Token

O que é: Uma pequena unidade usada pelo modelo para processar texto e outros dados.

Para quem tem menos de 30: A IA quebra conteúdo em pedaços para ler e escrever.

Para quem tem mais de 50: Lembra o telegrama: cada pedaço da mensagem influencia tamanho e custo.

Por que importa: Tokens afetam preço, velocidade e quantidade de conteúdo processado.

Não confunda: Um token não equivale sempre a uma palavra.

Janela de contexto

O que é: A quantidade de informação que o modelo consegue considerar durante uma tarefa.

Para quem tem menos de 30: É o tamanho da "mesa de trabalho" da IA.

Para quem tem mais de 50: É como o número de documentos que um assistente consegue manter abertos e relacionar ao mesmo tempo.

Por que importa: Contexto maior ajuda em projetos longos, mas não garante memória perfeita.

Não confunda: Contexto não é o mesmo que conhecimento permanente.

Agente de IA

O que é: Um sistema que recebe um objetivo, planeja etapas, usa ferramentas e tenta executar ações.

Para quem tem menos de 30: É como um personagem que recebe uma missão e pode usar itens do jogo.

Para quem tem mais de 50: Um chatbot responde; um agente se parece mais com um assistente que começa a trabalhar.

Por que importa: Agentes transformam IA de conversa em execução.

Não confunda: Agente não significa AGI, consciência ou autonomia ilimitada.

Frontier model

O que é: Um modelo que está entre os mais avançados disponíveis em determinado momento.

Para quem tem menos de 30: É um competidor na primeira fila da tecnologia atual.

Para quem tem mais de 50: É como um carro de Fórmula 1: representa o limite do momento e pode ser superado rapidamente.

Por que importa: Empresas frontier atraem capital, talentos e influência.

Não confunda: Frontier não significa perfeito, seguro ou consciente.

Open source AI

O que é: Um sistema oferecido com liberdades reais para usar, estudar, modificar e compartilhar.

Para quem tem menos de 30: É tecnologia que pode ser aberta e transformada pela comunidade.

Para quem tem mais de 50: É receber não apenas a máquina, mas meios suficientes para entender e adaptar seu funcionamento.

Por que importa: Abertura pode ampliar competição, soberania e auditoria.

Não confunda: Disponível para download não significa automaticamente open source.

Open weight

O que é: Um modelo cujos parâmetros finais treinados foram disponibilizados.

Para quem tem menos de 30: Você recebe o "cérebro pronto" para executar ou adaptar.

Para quem tem mais de 50: É como receber o motor montado, mas não necessariamente o projeto da fábrica, matérias-primas e processo completo.

Por que importa: Permite implantação local e customização.

Não confunda: Open weight e open source não são sinônimos.

Closed model

O que é: Um modelo controlado pela empresa e acessado por aplicativo ou API, sem distribuição de seus pesos completos.

Para quem tem menos de 30: Você usa o serviço, mas não recebe o sistema por dentro.

Para quem tem mais de 50: É como alugar uma máquina operada pelo fabricante.

Por que importa: Pode oferecer conveniência, suporte e segurança centralizada.

Não confunda: Fechado não significa automaticamente melhor nem pior.

API

O que é: Uma forma padronizada de um software conversar com outro.

Para quem tem menos de 30: É a porta pela qual um aplicativo chama uma IA.

Para quem tem mais de 50: É como um balcão de serviço com regras definidas para pedidos e respostas.

Por que importa: Permite colocar modelos dentro de empresas e produtos.

Não confunda: Usar API não significa que os dados ficam automaticamente privados.

Cloud / nuvem

O que é: Computadores e servidores remotos acessados pela internet.

Para quem tem menos de 30: É onde muitos aplicativos e arquivos realmente rodam.

Para quem tem mais de 50: Em vez de guardar tudo no computador da empresa, você aluga capacidade em data centers.

Por que importa: A nuvem tornou modelos poderosos acessíveis sem comprar toda a infraestrutura.

Não confunda: A nuvem é física: prédios, chips, energia, cabos e refrigeração.

Data center

O que é: Instalação física que concentra servidores, redes, energia e refrigeração.

Para quem tem menos de 30: É uma fábrica de computação.

Para quem tem mais de 50: É como uma grande usina que produz processamento em vez de um produto material.

Por que importa: A corrida de IA depende de energia, território, chips e água e refrigeração.

Não confunda: IA não vive em um espaço abstrato e sem custo ambiental.

Benchmark

O que é: Um teste usado para comparar modelos em tarefas definidas.

Para quem tem menos de 30: É uma prova aplicada a várias IAs.

Para quem tem mais de 50: É como comparar carros em pista, consumo e frenagem.

Por que importa: Ajuda a medir capacidade, mas não substitui teste no seu trabalho.

Não confunda: Primeiro lugar em uma prova não significa melhor em tudo.

Alucinação

O que é: Uma resposta falsa ou não sustentada apresentada de forma convincente.

Para quem tem menos de 30: É quando a IA inventa e parece ter certeza.

Para quem tem mais de 50: É como um funcionário que prefere improvisar uma resposta a admitir que não sabe.

Por que importa: Exige verificação em decisões importantes.

Não confunda: Texto elegante não é evidência.

Grounding

O que é: Ligação da resposta a fontes, documentos ou dados verificáveis.

Para quem tem menos de 30: É dar à IA o material real no qual ela deve se apoiar.

Para quem tem mais de 50: É pedir que o assistente trabalhe com os arquivos e registros da empresa, não apenas com memória geral.

Por que importa: Reduz erros e melhora rastreabilidade.

Não confunda: Ter uma citação não garante que ela foi interpretada corretamente.

Multimodal

O que é: Capacidade de trabalhar com mais de um tipo de informação, como texto, imagem, áudio e vídeo.

Para quem tem menos de 30: A IA pode ler, ver e ouvir diferentes mídias.

Para quem tem mais de 50: É como um profissional que recebe documento, fotografia, gravação e planilha na mesma tarefa.

Por que importa: Aumenta as aplicações em trabalho, ciência, educação e criação.

Não confunda: Multimodal não significa percepção humana completa.

AGI

O que é: Um conceito para inteligência artificial geral com capacidade ampla e adaptável comparável ou superior à humana em muitas tarefas; não existe consenso único sobre o teste definitivo.

Para quem tem menos de 30: Seria uma IA que aprende e trabalha em muitas áreas, não apenas uma ferramenta específica.

Para quem tem mais de 50: É a ideia de uma inteligência geral, em vez de uma máquina treinada para uma função estreita.

Por que importa: A busca por AGI orienta investimento e segurança.

Não confunda: Nenhuma empresa deve ser tratada como tendo alcançado AGI apenas porque usa a palavra.

ARR

O que é: Annual Recurring Revenue: receita recorrente anualizada com base no ritmo atual.

Para quem tem menos de 30: É projetar para um ano o que os contratos recorrentes estão gerando agora.

Para quem tem mais de 50: Se a empresa recebe US$ 10 milhões por mês de forma recorrente, pode falar em US$ 120 milhões de ARR.

Por que importa: É uma medida popular de crescimento de software.

Não confunda: ARR não é lucro, caixa recebido nem receita auditada de um ano completo.

Valuation

O que é: Estimativa de quanto uma empresa vale em uma rodada ou no mercado.

Para quem tem menos de 30: É o preço implícito da empresa, não o saldo bancário.

Para quem tem mais de 50: É semelhante ao valor atribuído a uma casa com base em uma transação, não ao dinheiro no cofre do proprietário.

Por que importa: Afeta capital, reputação e contratação.

Não confunda: Valuation não é receita nem garantia de sucesso.

IPO

O que é: Abertura de capital em que ações passam a ser oferecidas ao público em bolsa.

Para quem tem menos de 30: É quando investidores comuns podem comprar participação no mercado.

Para quem tem mais de 50: É a passagem de empresa privada para companhia aberta.

Por que importa: Pode financiar crescimento e criar liquidez, mas aumenta obrigações públicas.

Não confunda: Apresentar um S-1 não significa que a IPO tem data garantida.

Deepfake

O que é: Áudio, imagem ou vídeo sintético que imita pessoa ou evento de forma convincente.

Para quem tem menos de 30: É mídia falsa que pode parecer real.

Para quem tem mais de 50: É como uma montagem, mas com voz e movimento produzidos por IA em nível muito mais persuasivo.

Por que importa: Afeta fraude, reputação, política e confiança social.

Não confunda: Nem todo conteúdo gerado é malicioso; o risco está no engano.

Alfabetização em IA

O que é: A capacidade de compreender o que a IA faz, utilizá-la, verificar resultados e reconhecer riscos.

Para quem tem menos de 30: É mais do que saber abrir um chatbot; inclui fontes, privacidade, limites e responsabilidade.

Para quem tem mais de 50: É semelhante à alfabetização digital que passou a ser necessária quando computadores e internet entraram no trabalho.

Por que importa: Ela tende a se tornar uma competência básica em muitas profissões.

Não confunda: Alfabetização em IA não significa saber programar nem confiar em tudo que a ferramenta produz.

Contratação por competências

O que é: Uma forma de contratar que dá mais peso ao que a pessoa consegue demonstrar do que ao cargo anterior ou ao diploma isolado.

Para quem tem menos de 30: É provar o que você sabe fazer por meio de tarefas, portfólio e resultados.

Para quem tem mais de 50: É semelhante a pedir uma demonstração prática antes de contratar alguém para um trabalho especializado.

Por que importa: Currículos e textos podem ser produzidos por IA; evidência de competência se torna mais importante.

Não confunda: Remover exigências de diploma não elimina a necessidade de padrões rigorosos.

Work sample / amostra de trabalho

O que é: Uma tarefa curta e relevante usada para observar como o candidato trabalha.

Para quem tem menos de 30: É um teste parecido com o trabalho real, não uma prova genérica.

Para quem tem mais de 50: É como pedir a um fornecedor uma peça-amostra antes de aprovar uma produção maior.

Por que importa: Mostra raciocínio, qualidade, revisão e uso de ferramentas de maneira mais direta que um currículo.

Não confunda: A tarefa deve ser ética, limitada e não pode servir para obter trabalho gratuito.

Credencial

O que é: Um diploma, licença, certificado ou outro sinal formal de formação e autorização.

Para quem tem menos de 30: É o documento que comprova que você concluiu um programa ou foi autorizado a exercer algo.

Para quem tem mais de 50: É como uma carteira profissional: pode abrir a porta, mas não prova sozinha a qualidade de cada trabalho.

Por que importa: Algumas profissões exigem credenciais por lei; outras as usam apenas como filtro conveniente.

Não confunda: Credencial, competência e experiência são relacionadas, mas não idênticas.

Microcredencial

O que é: Um certificado curto e específico que comprova uma habilidade ou bloco de aprendizagem.

Para quem tem menos de 30: É uma formação menor do que uma faculdade, focada em uma competência.

Para quem tem mais de 50: É semelhante a um curso técnico ou certificação profissional com escopo bem definido.

Por que importa: Pode atualizar habilidades rapidamente, mas sua qualidade e reconhecimento variam muito.

Não confunda: Uma coleção de certificados não substitui automaticamente prática, base conceitual ou julgamento.

Cognitive offloading / terceirização cognitiva

O que é: Transferir para uma ferramenta parte do esforço de lembrar, calcular, escrever, navegar ou decidir.

Para quem tem menos de 30: É deixar o telefone lembrar caminhos e a IA preparar uma resposta.

Para quem tem mais de 50: É como usar calculadora, agenda ou secretária para não carregar tudo na cabeça.

Por que importa: Pode liberar capacidade para tarefas melhores, mas também impedir aprendizagem quando elimina o esforço necessário para formar uma habilidade.

Não confunda: Usar ferramenta não é preguiça; o risco aparece quando a pessoa perde a capacidade de compreender e conferir.

Avaliação autêntica

O que é: Uma avaliação que observa conhecimento em tarefas próximas da vida real, com processo, explicação e revisão.

Para quem tem menos de 30: É demonstrar que você sabe usar o que aprendeu, não apenas repetir uma resposta.

Para quem tem mais de 50: É semelhante a avaliar um profissional pelo trabalho que consegue executar e defender.

Por que importa: Na era da IA, provas orais, projetos, laboratórios e portfólios podem revelar melhor o aprendizado real.

Não confunda: Projeto bonito não prova domínio se ninguém verifica processo, autoria e compreensão.

Mais termos usados neste livro

Atlas Vivo da IA

O que é: A página digital que acompanha este livro e atualiza os fatos que mudam depressa, como modelos, preços, rankings e regras.

Para quem tem menos de 30: É a versão "sempre atualizada" das fotografias datadas do livro.

Para quem tem mais de 50: É como a errata e a nova edição de um guia, só que contínuas.

Por que importa: Na IA, um número correto em outubro pode estar velho em dezembro.

Não confunda: O livro ensina a pensar; o Atlas atualiza os fatos. Um não substitui o outro.

Automação e aumento

O que é: Duas formas de usar a IA. Na automação, a pessoa entrega a tarefa inteira; no aumento (augmentation), ela trabalha junto com a IA, revisando, iterando e decidindo.

Para quem tem menos de 30: Automação é "faz para mim"; aumento é "faz comigo".

Para quem tem mais de 50: É a diferença entre terceirizar um serviço e trabalhar lado a lado com um assistente competente.

Por que importa: O mesmo trabalho pode ser ampliado ou substituído dependendo de como a IA é usada.

Não confunda: Automatizar uma tarefa não é automatizar uma profissão inteira.

Capacidade computacional (compute)

O que é: O poder de processamento usado para treinar e rodar modelos de IA, medido em chips, servidores, energia e tempo de uso.

Para quem tem menos de 30: É o "combustível" que faz a IA funcionar.

Para quem tem mais de 50: É como a capacidade instalada de uma fábrica: sem ela, não há produção.

Por que importa: Quem controla chips, energia e data centers influencia quem consegue construir modelos de fronteira.

Não confunda: Ter mais capacidade computacional não garante, sozinho, um modelo melhor ou um negócio lucrativo.

Churn

O que é: A taxa de clientes que cancelam um serviço num período.

Para quem tem menos de 30: É quantas pessoas desistem da assinatura.

Para quem tem mais de 50: É a clientela que sai pela porta dos fundos enquanto você comemora quem entrou pela da frente.

Por que importa: Um ARR alto com churn alto é um balde furado.

Não confunda: Crescer em novos clientes não significa reter os antigos.

Companheiro virtual (AI companion)

O que é: Um aplicativo de IA desenhado para conversar como amigo, parceiro ou confidente, muitas vezes com personagem, memória e tom afetivo.

Para quem tem menos de 30: É o chatbot que "lembra de você" e conversa sobre a sua vida.

Para quem tem mais de 50: É como um correspondente sempre disponível, só que do outro lado há um software de uma empresa.

Por que importa: Pode aliviar a solidão, mas também criar dependência; vários governos passaram a discutir regras específicas, sobretudo para menores.

Não confunda: Um chatbot que conversa sobre emoções não é um psicólogo, mesmo quando parece um.

Conteúdo sintético

O que é: Texto, imagem, áudio ou vídeo criado ou alterado por inteligência artificial.

Para quem tem menos de 30: É tudo o que foi gerado ou editado por IA, do filtro à voz clonada.

Para quem tem mais de 50: É como uma fotomontagem, mas feita em segundos e muito mais difícil de perceber.

Por que importa: Afeta a confiança em provas, notícias e eleições; no Brasil, o TSE criou regras para o seu uso na propaganda eleitoral.[1]

Não confunda: Todo deepfake é conteúdo sintético, mas nem todo conteúdo sintético é deepfake ou enganoso.

Decisão automatizada

O que é: Uma decisão sobre uma pessoa tomada ou preparada por um sistema, como um preço, uma nota, uma aprovação, um bloqueio ou uma tarefa.

Para quem tem menos de 30: É quando o aplicativo decide e você só recebe o resultado.

Para quem tem mais de 50: É como receber uma carta do banco assinada por "o sistema", sem saber quem decidiu.

Por que importa: Novas regras, como a diretiva europeia sobre trabalho em plataformas, exigem supervisão humana, explicação e direito de revisão.[2]

Não confunda: Usar um aplicativo o dia todo não significa ter controle sobre as regras dele.

Diário de Fronteira

O que é: O caderno ou a nota no celular em que você registra, com data, as ferramentas que usa, os testes que faz e o que aprendeu.

Para quem tem menos de 30: É um log pessoal da sua evolução com IA.

Para quem tem mais de 50: É como o diário de bordo de quem navega: anota rota, clima e desvios.

Por que importa: Sem registro, você não sabe se melhorou, piorou ou apenas mudou de ferramenta.

Não confunda: Não é um diário íntimo; é um registro de trabalho e aprendizagem.

Ecossistema

O que é: Tudo o que vem junto com uma tecnologia: sistema operacional, loja, desenvolvedores, pagamentos, serviços e o hábito de quem usa.

Para quem tem menos de 30: É por que trocar de celular dá tanto trabalho.

Para quem tem mais de 50: É como um shopping inteiro em volta de uma loja âncora.

Por que importa: Quem controla o ecossistema costuma capturar mais valor do que quem faz só o aparelho.

Não confunda: Aparelho é o objeto; interface é a forma de conversar com ele; ecossistema é tudo o que vem junto.

Empresa de uma pessoa

O que é: Um negócio que opera com uma pessoa no centro e o resto do trabalho feito por ferramentas, agentes e parceiros externos.

Para quem tem menos de 30: É empreender sozinho com uma "equipe" de IA.

Para quem tem mais de 50: É o profissional autônomo que agora consegue entregar o que antes exigia um escritório inteiro.

Por que importa: Reduz o custo de tentar, mas não elimina a necessidade de clientes, qualidade e caixa.

Não confunda: Delegar tarefas à IA é possível; delegar a responsabilidade, não.

Engenharia de contexto

O que é: O cuidado com tudo o que a IA recebe além do pedido: documentos, histórico, instruções da ferramenta e dados.

Para quem tem menos de 30: É montar a "mochila" de informações que a IA leva para a tarefa.

Para quem tem mais de 50: É preparar a pasta completa antes de passar um caso a um colega.

Por que importa: Com modelos e agentes mais capazes, o que vai junto com o pedido pesa tanto quanto o pedido.

Não confunda: Prompt é o pedido que você escreve; contexto é tudo o que vai com ele.

Fato vivo

O que é: Uma informação verdadeira numa data, mas que tende a mudar logo, como preço, ranking, valuation ou versão de modelo.

Para quem tem menos de 30: É o dado com "prazo de validade".

Para quem tem mais de 50: É como a cotação do dólar: certa hoje, não necessariamente amanhã.

Por que importa: Neste livro, todo fato vivo vem com data e é atualizado no Atlas Vivo da IA.

Não confunda: Fato vivo não é boato; é um fato que precisa de data.

Fronteira irregular

O que é: A constatação de que a IA é excelente em algumas tarefas e falha em outras que parecem igualmente difíceis, ou até mais simples.

Para quem tem menos de 30: É o colega gênio em cálculo que esquece de salvar o arquivo.

Para quem tem mais de 50: É como um funcionário brilhante em algumas áreas que precisa de supervisão em outras.

Por que importa: Saber onde a IA acerta e onde erra é parte do trabalho de quem a usa.

Não confunda: Acertar uma tarefa difícil não garante acertar uma fácil.

GPU / chip de IA

O que é: Processadores capazes de fazer muitos cálculos em paralelo, usados para treinar e rodar modelos de IA.

Para quem tem menos de 30: São as placas que nasceram para os games e viraram o motor da IA.

Para quem tem mais de 50: São as máquinas-ferramenta da nova indústria.

Por que importa: A disputa por chips virou questão de empresas, de países e de segurança nacional.

Não confunda: O chip não é o modelo; é a infraestrutura em que o modelo roda.

IA generativa

O que é: Sistemas de inteligência artificial que produzem conteúdo novo, como texto, imagem, áudio, vídeo ou código, a partir de um pedido.

Para quem tem menos de 30: É a IA que cria, não apenas classifica.

Para quem tem mais de 50: É como passar do arquivo que encontra documentos para o assistente que escreve o documento.

Por que importa: Foi ela que levou a IA para o público a partir do fim de 2022.

Não confunda: Gerar conteúdo convincente não é o mesmo que gerar conteúdo verdadeiro.

Laboratório de fronteira (frontier lab)

O que é: A organização cuja missão central é construir os modelos mais avançados, como OpenAI, Anthropic ou xAI; o Google DeepMind é um laboratório de fronteira dentro de uma big tech.

Para quem tem menos de 30: É onde estão sendo feitos os modelos que todo mundo vai comparar.

Para quem tem mais de 50: É o centro de pesquisa e desenvolvimento que virou a própria empresa.

Por que importa: Atrai capital, talentos e influência política.

Não confunda: Big tech é a grande empresa estabelecida; laboratório de fronteira é a organização focada nos modelos mais avançados; startup de IA é a empresa nova que constrói produtos ou modelos próprios.

LLM (modelo de linguagem de grande porte)

O que é: Um modelo de IA treinado com enormes quantidades de texto para prever e produzir linguagem.

Para quem tem menos de 30: É o tipo de modelo por trás do ChatGPT, do Claude e do Gemini.

Para quem tem mais de 50: É um leitor incansável que aprendeu padrões de linguagem lendo uma biblioteca gigantesca.

Por que importa: É a base da maioria dos assistentes e agentes de hoje.

Não confunda: Prever bem a linguagem não é o mesmo que compreender como uma pessoa ou saber o que é verdade.

Marca d'água e rótulo

O que é: Sinais, visíveis ou invisíveis, que indicam que um conteúdo foi gerado ou alterado por IA.

Para quem tem menos de 30: É o selo "feito com IA".

Para quem tem mais de 50: É como o carimbo que identifica uma cópia.

Por que importa: Ajuda a identificar conteúdo sintético, mas quem sabe pode remover alguns desses sinais.

Não confunda: Rótulo não é proteção completa: um conteúdo pode causar dano mesmo avisando que foi feito por IA.

Margem bruta

O que é: O que sobra da receita depois do custo direto de entregar o produto, como os modelos de IA alugados.

Para quem tem menos de 30: É quanto fica de cada venda antes das outras despesas.

Para quem tem mais de 50: É o velho cálculo de preço de venda menos custo da mercadoria.

Por que importa: Muitas empresas de IA crescem rápido, mas pagam caro pelos modelos que usam.

Não confunda: Margem bruta não é lucro; ainda faltam salários, marketing, impostos e outras despesas.

Modelo de raciocínio

O que é: Um modelo que usa mais tempo e mais processamento para produzir etapas intermediárias antes de dar a resposta final.

Para quem tem menos de 30: É a IA que "pensa antes de responder".

Para quem tem mais de 50: É como o profissional que faz rascunho e confere a conta antes de entregar.

Por que importa: Melhora o desempenho em matemática, programação e ciência, mas pode custar mais por tarefa.

Não confunda: Mostrar etapas não garante que o raciocínio esteja certo, nem que seja o processo interno real do modelo.

Parâmetros / pesos

O que é: Os números ajustados durante o treinamento que guardam o que o modelo aprendeu.

Para quem tem menos de 30: São as "configurações internas" do cérebro da IA.

Para quem tem mais de 50: São como a regulagem final de um motor depois de anos de testes.

Por que importa: Quem libera os pesos permite que outros rodem e adaptem o modelo; o número de parâmetros dá uma ideia do tamanho do modelo.

Não confunda: Mais parâmetros não significam automaticamente um modelo melhor.

Preço por token

O que é: O valor cobrado por cada pedaço de texto que o modelo lê ou escreve, geralmente informado por milhão de tokens.

Para quem tem menos de 30: É a tarifa da IA, como o preço do dado móvel.

Para quem tem mais de 50: É como a tarifa por minuto da telefonia antiga.

Por que importa: A queda do preço por capacidade tornou possível usar IA em tarefas que antes não compensavam.

Não confunda: Preço por token não é custo por tarefa: modelos de raciocínio podem cobrar menos por token e usar muito mais tokens.

Ranking

O que é: Uma classificação de modelos construída a partir de uma ou mais provas ou da preferência de usuários.

Para quem tem menos de 30: É a tabela do campeonato das IAs.

Para quem tem mais de 50: É como o ranking de vendas de carros: informa, mas não escolhe por você.

Por que importa: Ajuda a saber por onde começar a testar.

Não confunda: Benchmark é a prova; ranking é a classificação; alegação da empresa é o resultado que ela escolheu divulgar.

Receita

O que é: O dinheiro que a empresa efetivamente faturou com vendas num período fechado.

Para quem tem menos de 30: É o que entrou de verdade, não o que se projeta.

Para quem tem mais de 50: É o faturamento do ano, conferido pela contabilidade.

Por que importa: Receita auditada é a base mais segura para comparar empresas.

Não confunda: Receita não é ARR (o ritmo atual anualizado), nem lucro, nem valuation.

Regra 1-1-1-30

O que é: A regra proposta no Capítulo 4: uma IA principal, uma especializada apenas se necessária, um processo real e uma revisão a cada 30 dias.

Para quem tem menos de 30: É foco em vez de testar todo app novo.

Para quem tem mais de 50: É começar pequeno, medir e só depois ampliar.

Por que importa: Transforma curiosidade em resultado sem se perder no excesso de ferramentas.

Não confunda: Não limita a curiosidade; dá a ela um lugar para produzir resultado.

Rodada de investimento

O que é: Uma captação em que investidores colocam dinheiro numa empresa em troca de participação, a um valuation negociado.

Para quem tem menos de 30: É quando a startup "levanta" dinheiro.

Para quem tem mais de 50: É como trazer novos sócios capitalistas para o negócio.

Por que importa: Define quanto capital a empresa tem para crescer e quanto cada sócio passa a ter.

Não confunda: Capital comprometido não é dinheiro já recebido: pode chegar em parcelas e com condições.

S-1

O que é: O documento de registro que uma empresa entrega à SEC, a comissão de valores mobiliários dos Estados Unidos, para vender ações ao público.

Para quem tem menos de 30: É o "raio-x" oficial que a empresa mostra antes de entrar na bolsa.

Para quem tem mais de 50: É o prospecto da oferta, com números, riscos e processos judiciais.

Por que importa: Quando se torna público, revela dados que a empresa antes não divulgava.

Não confunda: S-1 confidencial não é IPO marcada: é a opção de abrir o capital, se e quando a empresa quiser.

Soberania em IA

O que é: A capacidade de um país ou região controlar parte da sua própria infraestrutura de IA: dados, modelos, chips, nuvem, energia e regras.

Para quem tem menos de 30: É não depender totalmente de outro país para usar IA.

Para quem tem mais de 50: É como ter refinaria própria em vez de só importar combustível.

Por que importa: Afeta segurança, economia, língua, cultura e poder de negociação.

Não confunda: Soberania não significa fazer tudo sozinho nem fechar as portas a parceiros.

Superinteligência

O que é: Um conceito para uma IA que superaria amplamente os humanos em praticamente todas as tarefas cognitivas.

Para quem tem menos de 30: É a ideia de uma IA mais capaz do que qualquer pessoa em quase tudo.

Para quem tem mais de 50: É o "e se" mais discutido por cientistas, empresas e governos.

Por que importa: Algumas empresas já usam a palavra no nome de laboratórios e na disputa por talentos.

Não confunda: Usar a palavra não significa tê-la alcançado; não há consenso sobre se ou quando isso acontecerá.

Treinamento e inferência

O que é: Treinamento é a fase em que o modelo aprende com grandes volumes de dados; inferência é o uso do modelo já treinado para responder a pedidos.

Para quem tem menos de 30: Treinar é estudar para a prova; inferência é fazer a prova.

Para quem tem mais de 50: Treinamento é construir a fábrica; inferência é cada produto que sai dela.

Por que importa: O treinamento de fronteira custa bilhões; o custo da inferência é o que você sente no preço do uso.

Não confunda: Um modelo não aprende permanentemente com cada conversa só porque você o usa.

Viés algorítmico

O que é: Um erro sistemático de um sistema de IA que prejudica mais alguns grupos de pessoas do que outros.

Para quem tem menos de 30: É quando a IA funciona bem para uns e mal para outros.

Para quem tem mais de 50: É como uma régua torta: mede todos, mas não do mesmo jeito.

Por que importa: Em 2018, sistemas comerciais de reconhecimento facial erraram muito mais com mulheres de pele escura do que com homens de pele clara.[3]

Não confunda: Um sistema pode ter alta precisão média e, ainda assim, ser injusto com um grupo.

O glossário reverso: o mundo que os mais jovens não precisaram aprender

Lista telefônica. Um livro físico com nomes, endereços e telefones; lembra que "buscar" já significou procurar alfabeticamente sem atualização instantânea.

Telefone fixo. Um aparelho preso a um lugar; ajuda a explicar por que ligar significava encontrar uma casa ou empresa, não uma pessoa em qualquer lugar.

Secretária eletrônica. Uma máquina que gravava recados quando ninguém atendia; antecedeu voicemail, texto e mensagens assíncronas.

Fax. Documento enviado por linha telefônica; mostra como transmitir uma página já foi uma inovação empresarial.

Pager. Dispositivo que recebia números ou mensagens curtas; criou urgência móvel antes do celular.

Internet discada. Conexão lenta que ocupava a linha telefônica; lembra que estar online já era uma atividade temporária.

Disquete. Mídia física de armazenamento com capacidade minúscula para padrões atuais; símbolo de quando transportar dados exigia um objeto.

Enciclopédia. Coleção de livros consultada para pesquisa geral; mostra o valor que síntese, curadoria e autoridade tinham antes da busca.

Mapa de papel. Ferramenta de navegação sem localização automática; exigia planejar, interpretar e lembrar caminhos.

Nokia e BlackBerry. Líderes de telefonia móvel sob as regras anteriores ao smartphone moderno; metáfora para empresas excelentes que enfrentam mudança de interface.

IBM Aptiva. Linha de computadores pessoais da IBM; parte da minha experiência na democratização do PC no varejo brasileiro.

Loja de aplicativos inexistente. Antes do iPhone, telefone e software não formavam o mesmo mercado de desenvolvedores, pagamentos e atualizações que hoje parece natural.

Notas

1. Tribunal Superior Eleitoral. Resolução nº 23.755, de 2 de março de 2026 (altera a Resolução nº 23.610/2019, art. 9º-B, § 3º-A). https://www.tse.jus.br/legislacao/compilada/res/2026/resolucao-no-23-755-de-2-de-marco-de-2026; TSE. "Por Dentro das Eleições: conheça as regras sobre uso de IA na campanha eleitoral de 2026." Abr. 2026. https://www.tse.jus.br/comunicacao/noticias/2026/Abril/por-dentro-das-eleicoes-conheca-as-regras-sobre-uso-de-ia-na-campanha-eleitoral-de-2026

2. União Europeia. Diretiva (UE) 2024/2831 do Parlamento Europeu e do Conselho, relativa à melhoria das condições de trabalho em plataformas digitais. *Jornal Oficial da União Europeia*, 2024 (transposição até 2 dez. 2026). https://eur-lex.europa.eu/eli/dir/2024/2831/oj

3. Buolamwini, Joy; Gebru, Timnit. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." *Proceedings of Machine Learning Research*, v. 81, p. 77–91, 2018 (Conference on Fairness, Accountability and Transparency). https://proceedings.mlr.press/v81/buolamwini18a.html; MIT News, 12 fev. 2018. https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212