



meio
perspectiva:

IA

2024



Índice

- 5__ A corrida da inteligência artificial
- 12__ O que pensam nossos leitores?
- 15__ Arte artificial
- 21__ Qual a ferramenta certa para você?
- 35__ A última fronteira
- 41__ A IA super-humana
- 44__ Aprendizados na máquina
- 46__ Glossário



Pedro Doria

A corrida da inteligência artificial

Para que a IA consiga transformar o mundo como a internet o fez, há três aspectos a se observar: o geopolítico, o desenvolvimento tecnológico e os modelos de negócio

Estamos no ano dois da corrida da inteligência artificial. A internet demorou 20 anos para se tornar amplamente distribuída, via smartphones na mão, às vezes parece, de todo mundo. A conta da maioria dos analistas dá que para IA o processo será na metade do tempo. Por conta da nova tecnologia, o poder computacional total no mundo vem dobrando a cada dois anos. De todo esse arsenal de processamento, aquele voltado só para treinar modelos inteligentes vem decuplicando a cada ano, ou seja, aumentando em dez vezes. E isso ocorre porque a corrida não é metafórica: ela é real, uma disputa de fato que passa por algumas categorias diferentes.

EUA VS CHINA

A face mais evidente é a geopolítica, a briga entre China e Estados Unidos. O placar do momento é ambíguo, mas revela muito. No top 20 das entidades que mais publicaram patentes relacionadas a IA, 13 são chinesas e apenas quatro, americanas — Japão tem duas e Coreia do Sul, uma. O top 4 é todo chinês. A líder americana, IBM, está em quinto nessa lista, seguindo a Academia Chinesa de Ciências, que vem após o Baidu, uma seguradora chamada Ping e, na primeira posição, a Tencent. Mesmo que olhemos para além do ranking, os EUA estão mal. Registraram 21% do total de patentes em inteligência artificial. A China, 61%, até 2023. Os números parecem contar uma história, a de que em produção de conhecimento há uma disputa muito desigual.

A métrica engana. No ano de 2023, os Estados Unidos desenvolveram 61 modelos de inteligência artificial, a União Europeia desenvolveu 21, e a China, 15. Em tecnologia, a ideia que o papel aceita no registro de patentes só tem valor quando se torna produto. E os chineses têm profunda dificuldade de botar em prática o que imaginam. Existe uma razão para que essa inversão ocorra: modelo. Nesse caso, não um modelo de IA, mas o modelo de país.

Para que um modelo de inteligência artificial nasça, é preciso acesso a três recursos em vasta quantidade. O primeiro, evidentemente, é poder computacional. São necessários muitos computadores — a capacidade de processamento de dados não vem decuplicando anualmente no setor à toa. A China tem computadores, mas os limites impostos pelos EUA à exportação dos microchips adequados dificultam hoje a montagem dos parques necessários de treinamento. Ainda assim, esse é um problema que em alguns anos, com a formidável indústria chinesa, é possível resolver.

“ A criação de um modelo de inteligência artificial demanda acesso a três recursos em vasta quantidade. O primeiro, evidentemente, é poder computacional; o segundo, dinheiro; e o terceiro são dados ”

O segundo recurso é dinheiro. No ano de 2022, os americanos investiram US\$ 47 bilhões no desenvolvimento de IA. A China, US\$ 13 bilhões. A dificuldade não se dá, apenas, na profundidade de cada bolso. Os chineses têm dinheiro. Mas onde e como investir? Boa parte do investimento americano é privado, feito em empresas igualmente privadas, que competem entre si. São todas capitaneadas por cientistas na casa dos 30, de olho em construir o futuro, revolucionar o mundo, fazer fama e, ao fim, construir imensas fortunas pessoais. O produto dessa sopa de incentivos é um conjunto pulverizado de startups originais, com grande potencial, todas com tecnologia na rua em forma de produto. Não bastasse, o risco é distribuído entre investidores diferentes e empresas distintas.

Enquanto isso, nos últimos cinco anos o governo chinês promoveu uma grande ofensiva contra os principais líderes de empresas de tecnologia, cerceando seus espaços e impondo mais controle sobre o que fazem. A principal vítima foi justamente a estrela nascente, símbolo do Vale do Silício chinês, Jack Ma, o carismático fundador do grupo Alibaba. De frequentador das conferências internacionais, silenciou-se. O recado foi claro para todos os fundadores, empresários e investidores: Pequim não quer que se sobressaiam demais ou tenham ideias próprias. Podem enriquecer, não podem achar que são ocidentais. A consequência é menor estímulo para inovar e menos diversidade. Mesmo com dinheiro para investir, as opções são rarefeitas.

E há o terceiro recurso, um dos mais complicados de encontrar: dados. Modelos de inteligência artificial são treinados com vastas quantidades de dados — dados que precisam ter qualidade, diversidade e ser organizados. Um regime que controla informação, que cerceia acesso à internet aberta, impõe limites que uma democracia simplesmente não tem. Só a camada de burocracia para ter acesso às devidas permissões já retarda o jogo. Os controles necessários para quem pode acessar o quê, sempre sob a vigília do Partido Comunista, já diminui ainda mais o número de empresas que poderiam criar algo.

No ano dois da Era da Inteligência Artificial, ao menos na largada, o modelo de desenvolvimento de tecnologia dos EUA oferece uma nítida vantagem competitiva quando comparado ao da China. E isso valeria mesmo que não houvesse barreira comercial com microchips.

A GUERRA DOS CHIPS

Pois a corrida dos chips é outra disputa desigual. No último 19 de junho, a Nvidia ultrapassou a Microsoft e se tornou a companhia de maior valor de mercado do mundo. Alcançou uma avaliação de US\$ 3,34 trilhões. Não manteve a posição, mas vem flutuando desde então com Microsoft e Apple na disputa pelo top 3. E, ainda assim, não é uma empresa imediatamente conhecida por todos.

A Nvidia fabrica microchips. Foi fundada em 1993, no Vale do Silício, por um trio de engenheiros com uma ideia. Os computadores estavam ganhando habilidades gráficas, abandonavam a tela preta cheia de letrinhas por uma colorida, com pastas e a seta do mouse. Para que os videogames se sofisticassem, no entanto, seria preciso que aquelas máquinas pudessem processar esses gráficos mais rapidamente. A ideia dos fundadores foi criar uma unidade de processamento exclusiva para isso. O computador teria um segundo chip. E o invento colou, o que permitiu que videogames dessem um salto de sofisticação na virada para o século 21.

A partir de 2012, o CEO, Jensen Huang, fez uma aposta concentrada em inteligência artificial. Ele queria criar chips aptos particularmente para processar modelos. A Mizuho Securities estima que hoje, nos parques de computadores que treinam modelos de inteligência artificial, algo entre 70% e 95% dos chips responsáveis pelo trabalho pesado são da Nvidia. A aposta deu muito certo.

Mas isso quer dizer, também, que a companhia é o maior gargalo do treinamento de IA que existe. Se não há mais computadores criando mais rápido novos modelos é porque a empresa de Huang não consegue produzi-los em maior quantidade. Também aí, evidentemente, há uma corrida. Concorrentes como AMD, Intel e outras estão produzindo chips de aceleração de IA. O virtual monopólio da Nvidia, inevitavelmente, cederá. Será um momento de avanço maior.

“ A última corrida em curso ainda passa despercebida para boa parte dos que deveriam estar na disputa. É a corrida da adaptação ao trabalho num mundo com IAs ”

MODELOS DE NEGÓCIO

Esta, evidentemente, não é a única corrida em curso. Há uma disputa a respeito do modelo de negócios que corre em duas raias diferentes. Uma é a briga entre IA livre, de código aberto, e IA proprietária. Outra é sobre o acesso via assinatura versus o gratuito.

O debate entre código aberto, software livre, de um lado, e o pago do outro é tão antigo quanto a própria indústria do software. Mas ele se transformou com o crescimento da internet. As grandes plataformas digitais escolheram ser abertas e se financiar por publicidade, ao invés de cobrar pelo uso. Sejam os sites de busca, ou as redes sociais, bibliotecas de vídeo como o YouTube, todas fazem dinheiro da mesma forma. Para quem acessa, são gratuitas. Em compensação, os dados pessoais são sugados e transformados em renda via propaganda. Com o porte que esses negócios adquiriram, esse formato provou-se vencedor.

Um dos modelos de maior sucesso de geração de imagens, o Stable Diffusion, é distribuído em código aberto. Não quer dizer que ele seja público — não, o modelo pertence à Stability AI, uma companhia britânica, que no entanto permite a qualquer um que o instale em seus servidores fazer uso livre dele. A condição é de que quaisquer mudanças implementadas sejam compartilhadas. Se constrói uma grande comunidade, que terminará tendo um produto mais sólido, com maior capacidade do que os da concorrência. Ao menos essa é a lógica. Ainda há um debate interno sobre como viabilizar financeiramente o negócio — mas isso é comum em novos mercados digitais.

A distribuição por código aberto oferece vantagens. A existência de opções abertas impede as distorções naturais de mercados muito concentrados, que são um risco no mundo de IA. Haverá sempre uma opção livre. A transparência de como a tecnologia funciona é também maior, pois mais gente tem acesso a sua estrutura. Além disso, democratiza acesso. Qualquer um pode instalar Stable Diffusion e criar as imagens que desejar, sem quaisquer limites, e a um custo muito baixo.

Tem, porém, desvantagens. Inteligência artificial traz riscos como a criação de deep fakes, imagens falsas que parecem em todo genuínas. Produtos comerciais rivais do Stable Diffusion, como Dall-E da OpenAI ou Midjourney, têm filtros que impedem a produção de fotografias

falsas com rostos de pessoas reais, bloqueiam conteúdo pornográfico, ultraviolento.

Essa briga é pesada. As big techs vêm fazendo lobby por regulamentação de inteligência artificial, alertando para os riscos de segurança. Em parte, esse lobby que poderia ser incoerente é em verdade um pedido de reserva de mercado: impor limites tais que impeçam as opções livres de existirem. Mas, para além da esperteza, o debate sobre dar livre acesso a ferramentas de IA, acesso para todo mundo e em qualquer lugar, jamais foi tido com profundidade ou levado a público. Já vivemos num ambiente em que discriminação e desinformação vêm crescendo, um ambiente em que a política é mais instável. O risco de esse ambiente piorar, escancarando acesso às ferramentas sem quaisquer filtros, não é trivial.

Neste momento, as ferramentas fechadas têm performance bastante superior: 24,2% acima das ferramentas abertas, segundo o AI Index 2024 Annual Report, da Universidade de Stanford.

A outra discussão é ainda incipiente, mas no mercado há uma preferência. O caminho para fazer dinheiro pode ser o das assinaturas, preferido da indústria de software profissional, ou o do acesso gratuito via publicidade, como das plataformas de internet. A escolha do modelo de assinatura tem razão de ser. Segurança de dados. A sociedade compreende cada vez mais que publicidade digital tem um custo alto, o de ver circular nossas informações digitais, que se espalham pela internet manipuladas de todas as formas. Com a assinatura, vem junto a garantia de que o privado seguirá sendo. Inteligência artificial é capaz de tirar conclusões sensíveis a respeito da vida financeira, saúde, sexualidade, relacionamentos. Esse poder só vai ampliar. Parece razoável que o modelo por assinatura se imponha.

Mas ele tem uma desvantagem muito importante: custa dinheiro. Ou seja, é para quem pode pagar, não basta ter acesso a internet. E, nesse caso, pode ser uma opção que aumenta a desigualdade que já existe, separando os que têm a vida facilitada por IA e os que não podem ter.

Esses são debates que não podem ser limitados à esfera das companhias que desenvolvem a tecnologia. Precisam envolver os legisladores. Precisam chegar, principalmente, à sociedade.

E O TRABALHO

A última corrida em curso, igualmente disparada em 2022, ainda passa despercebida para boa parte dos que deveriam estar na disputa. É a corrida da adaptação ao trabalho num mundo com IA.

A Microsoft, que por conta de seu investimento na OpenAI (ChatGPT) se tornou uma das companhias mais relevantes nesse ambiente, foca todo seu marketing em uma única palavra: Copilot. Todos seus produtos de inteligência artificial vêm com a marca que, em seu nome, busca tranquilizar as pessoas. É um copiloto. Uma ferramenta para ajudar, para agilizar, não para substituir ninguém. Os executivos nesta indústria, em suas entrevistas e falas públicas, batem na mesma tecla. IA não substitui — facilita.

O último relatório da McKinsey reafirma esse ponto. Embora possa automatizar algumas tarefas, seu maior potencial é o de aumentar a capacidade humana. Substituir nas tarefas mais tediosas, permitindo mais tempo para foco no que é criativo ou estratégico. O estudo da consultoria tira suas conclusões das entrevistas feitas com quem trabalha em inúmeros setores e já usa a tecnologia, assim como as conversas com os desenvolvedores. Não é apenas um chute. O impacto será percebido em todos os níveis do mercado — da checagem de estoque ao trabalho intelectual, do caixa ao CEO.

Mesmo que inteligência artificial crie riqueza, aumente o PIB, sem provocar uma onda de desemprego, adaptação é inevitável. Uma nova série de habilidades precisará ser treinada. São, por exemplo, habilidades cognitivas: pensar de forma crítica, criativa, habilidade de identificar problemas e resolvê-los. Processar informação. Parecem lugares comuns, mas não são. O trabalho mental que frequentemente fazemos de forma quase automatizada poderá ser feito por algoritmos. Isto quer dizer que sobra aquilo que IA não resolve. Conferir, refletir, ligar pontos. De certa forma, é mais trabalho. Justamente o trabalho mais difícil. Também mais prazeroso. Quem não tem o hábito de produzir assim precisa criar. E treinar.

Entramos num mundo em que fazer perguntas bem, com precisão, se tornará uma arte sofisticada: como extrair da IA a melhor resposta para o problema dado.

Em empresas demais, pessoas em quantidade acostumaram-se com trabalhos repetitivos. Suas vidas estão para mudar.

***Pedro Doria** é diretor de jornalismo do Meio*

O que pensam nossos leitores?

Conhecer nossa audiência e produzir jornalismo qualificado que expanda seu entendimento sobre os temas mais relevantes do momento está entre as missões do Meio.

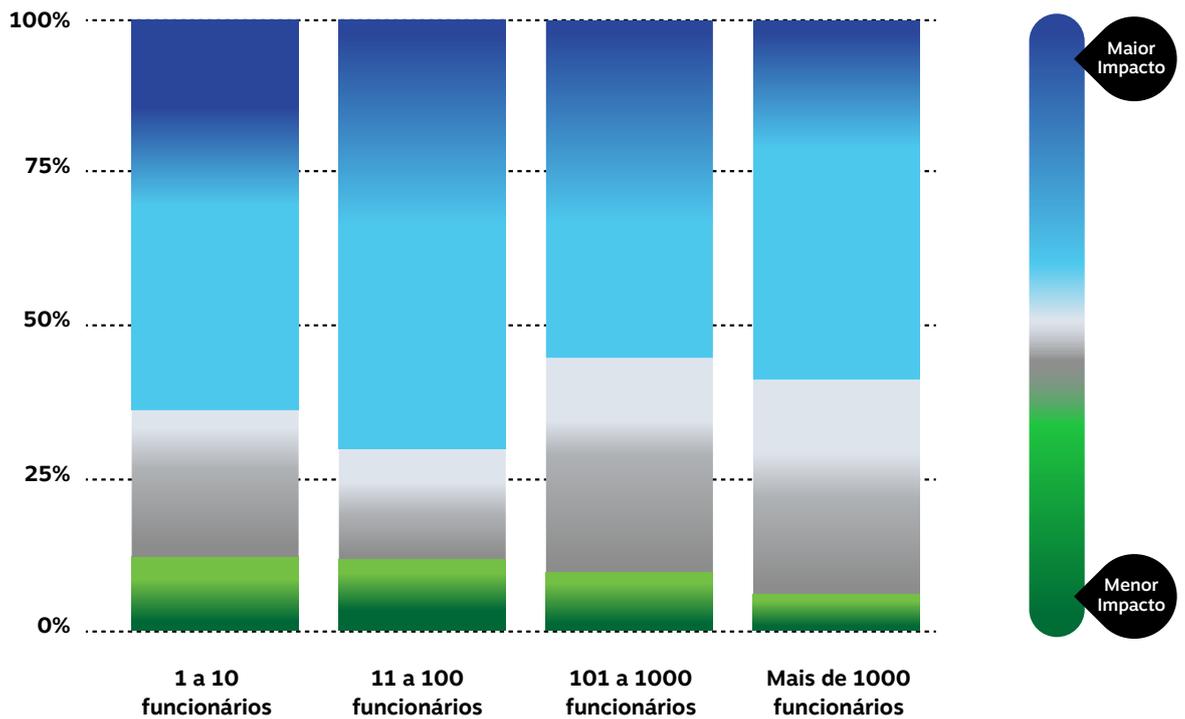
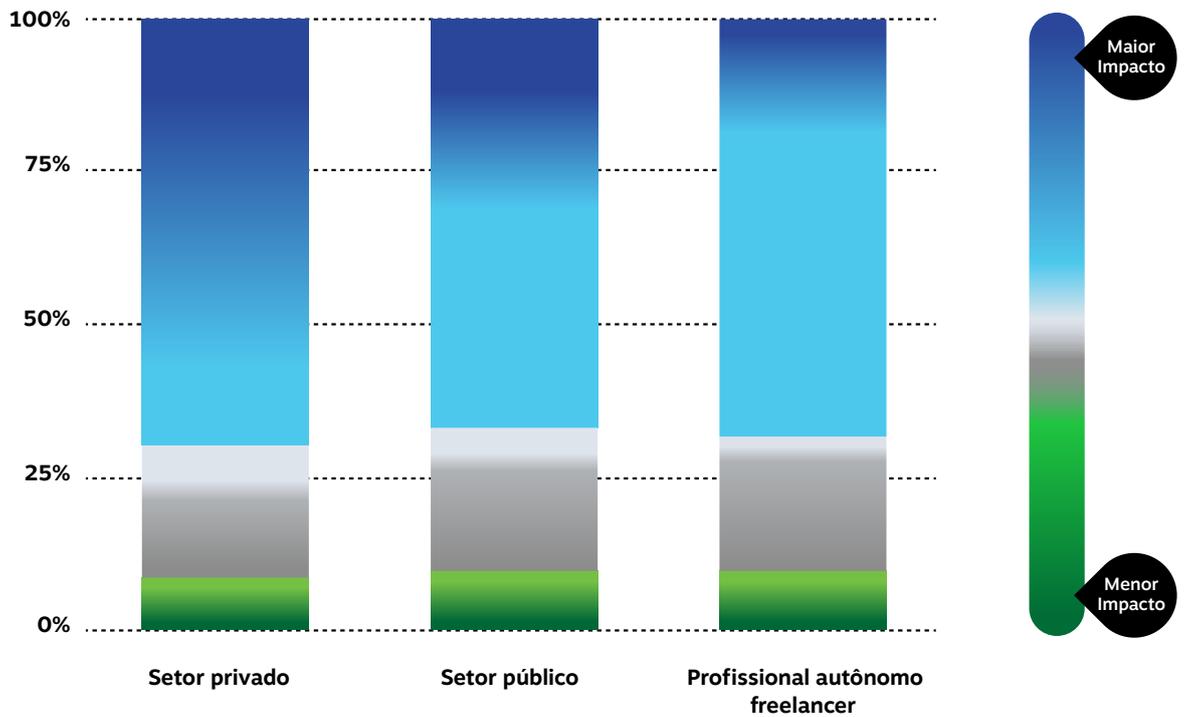
No início de julho, realizamos uma pesquisa abrangente com a base de leitores do Meio. Contamos com a participação de 544 respondentes. Traçamos um panorama de como as tecnologias de IA generativa estão sendo adotadas, quais ferramentas são mais utilizadas e os impactos percebidos no ambiente corporativo.

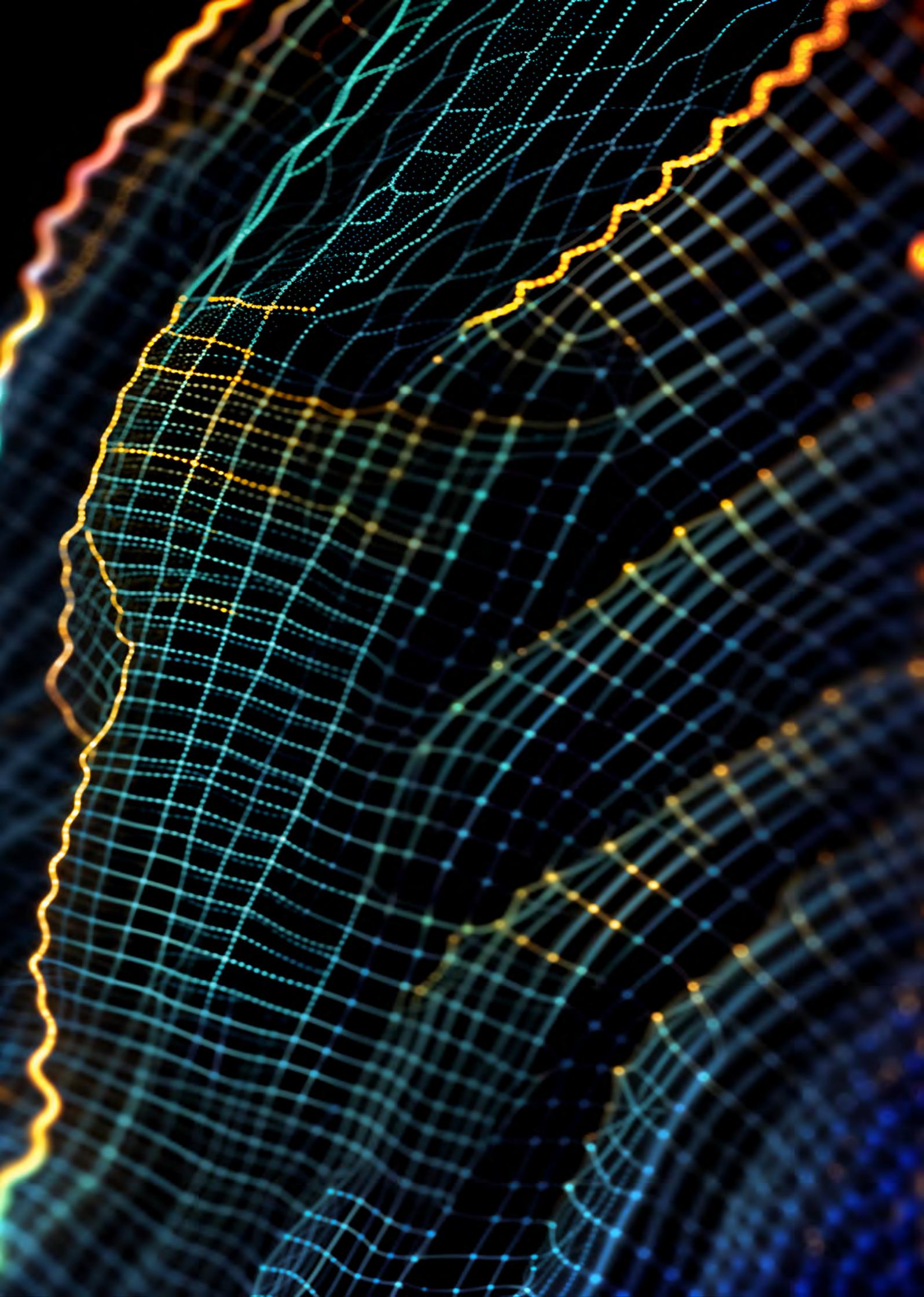
Um dos dados revelados pela pesquisa foi o de que uma parcela significativa dos participantes (36%) já paga por serviços de IA.

A maioria dos respondentes tem pós-graduação (61%) ou ensino superior completo (30%). As áreas de atuação são diversas, refletindo um público altamente qualificado.

Mais informações sobre como nossos leitores estão percebendo e usando IA estão em gráficos distribuídos nesta edição.

Como você avalia o impacto da IA na sua área de atuação?





Nelson Porto

Arte artificial

Para navegar no espaço latente da inteligência artificial ainda é preciso uma pessoa, uma intenção, um destino, uma consciência

Não é coincidência que as palavras “arte” e “artificial” compartilhem em sua origem o mesmo radical etimológico “ar”. Em seu sentido primeiro, ar- significava combinar ou unir duas coisas para criar algo novo. Na evolução da nossa linguagem, se combinamos coisas com a intenção de expressar sentimentos, isso é arte; se a intenção é simular ou criar um efeito, é um artifício, uma técnica, uma forma de se conseguir algo por meios não naturais. Como os fogos de artifício.

Já a expressão “inteligência artificial”, que aparece em duas de cada três frases que escutamos hoje, tem sido usada de maneira tão solta e carregada de sentidos distorcidos que dificilmente significará a mesma coisa para duas pessoas distintas. Contaminada por altas doses de ficção científica, embebida em narrativas apocalípticas

e ilustrada por robôs fora de controle, não à toa muitos na área preferem falar sobre aprendizado de máquina (*machine learning*), redes neurais e modelos generativos.

De fato, a tecnologia não é nova, mas seu desenvolvimento extremamente rápido nos últimos quatro anos pegou boa parte do mundo de surpresa, e essa velocidade e incompreensão sobre como ela funciona vêm gerando apreensão, insegurança e resistência em boa parte da população.

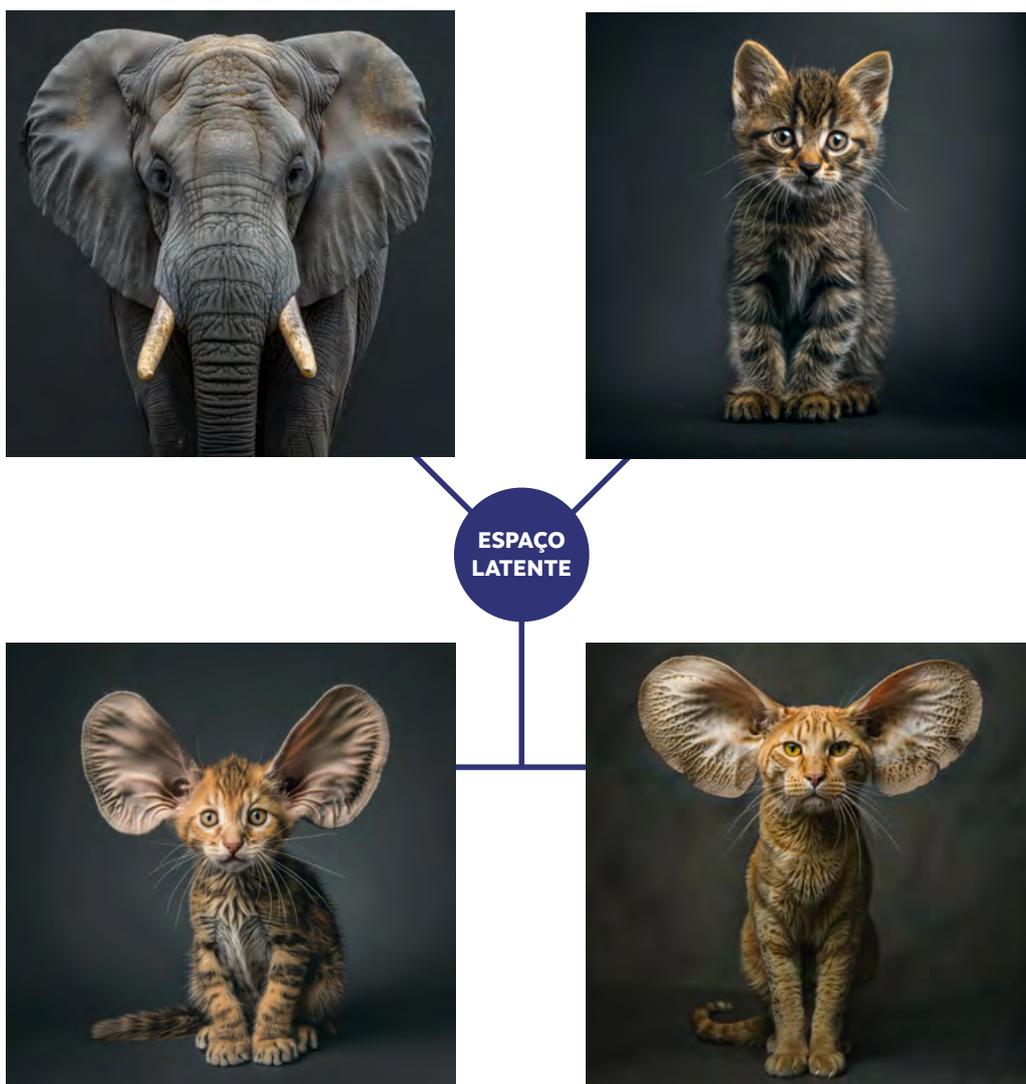
Talvez por isso seja realmente melhor falarmos sobre esse aprendizado, um conceito bem mais familiar cujo sentido é compreendido de forma mais ampla e uniforme. E como é que nós, humanos, aprendemos algo novo? Como as crianças aprendem a andar, a falar, a desenhar? Primatas que somos, aprendemos por observação e imitação. Aprender é, por definição, se apropriar de algo — seja de uma técnica, uma palavra, ou um conceito.

Para aprender, precisamos de treinamento: de exposição, repetição, erros e tentativas de reprodução, até chegarmos a um bom resultado. Só então podemos dizer que aprendemos algo. Ao combinar e misturar essas coisas que aprendemos treinando, muitas vezes acabamos por criar algo novo, que ainda não havia sido feito — e temos tanto orgulho de quando fazemos isso que às vezes até chamamos de arte.

No aprendizado de máquina não é diferente. As chamadas redes neurais foram estruturadas de forma a simular uma rede de neurônios humanos, em que cada ponto da rede guarda um pedaço da informação. Ao treinar uma rede neural com uma base de imagens, ela está observando e aprendendo infinitos padrões e conceitos que ela é capaz de identificar nessas imagens, sem que esteja copiando essas imagens como alguns tendem a imaginar.

Isso quer dizer que, ao observar em treinamentos repetidos milhares de imagens de elefantes e gatos, por exemplo, a máquina acaba por compreender os conceitos de gato, elefante e também os detalhes, como os pêlos e as orelhas de gato, ou a pele e as orelhas do elefante. Isso permitirá que ela crie infinitos “gatos com orelhas de elefante” após o treinamento. Como ela entende todos os conceitos envolvidos, ela pode recombina-los para criar algo novo, que não estava na base de imagens original.

“ Dependendo do tamanho e da qualidade da informação, esse espaço latente guarda tesouros infinitos ”



No espaço latente, existem pontos em que gatos têm orelhas de elefante

ESPAÇO LATENTE

O espaço latente é esse espaço abstrato do aprendizado de máquina, em que a rede neural guarda infinitos conceitos e padrões aprendidos em seu treinamento. É um verdadeiro oceano de possibilidades, pois nesse espaço é possível navegar por inúmeros locais que estão entre inúmeros pontos, e representam novas combinações de padrões e conceitos que antes não eram possíveis ou visíveis para nós.

Dependendo do tamanho e da qualidade da informação, esse espaço latente guarda tesouros infinitos, e a inteligência artificial nada mais é do que a ferramenta de que dispomos para navegar e explorar esse espaço do conhecimento e da criação humana. No entanto, creio estarmos muito distantes do ponto em que essa ferramenta deixe de precisar de um ser humano para funcionar. No fundo, esses modelos são apenas conjuntos de pesos estatísticos, coleções de números organizados de uma forma brilhante, porém estéril, desprovida de consciência, vontade e espírito.

Esse processo de associação de conceitos foi modelado a partir do funcionamento do cérebro humano, e acaba sendo muito similar ao que acontece em algumas etapas dos nossos processos criativos. Mas o desenvolvimento dessas capacidades nos humanos não aconteceu do dia para a noite. Pelo contrário, ele se deu de forma orgânica, sobre uma base construída desde o início da evolução da vida na Terra em 4,5 bilhões de anos — e há muitas camadas e níveis de inteligência envolvidos e desenvolvidos antes de se chegar ao estágio da criatividade humana.

O QUE É QUE IA NÃO TEM?

Essa base evolutiva que os humanos trazem do nascimento, de camadas de consciência e percepção sendo treinadas exaustivamente por meio dos sentidos, a IA não tem. E sendo treinada exclusivamente em textos, sons e imagens, como as LLMs (Large Language Models ou modelos de linguagem de grande escala, em português), ela jamais terá. Um LLM dificilmente conseguirá construir um modelo funcional do mundo, e isso se torna aparente quando se pede que as IAs mais avançadas da atualidade, como o Chat-GPT4, solucionem problemas simples, e elas alucinam.

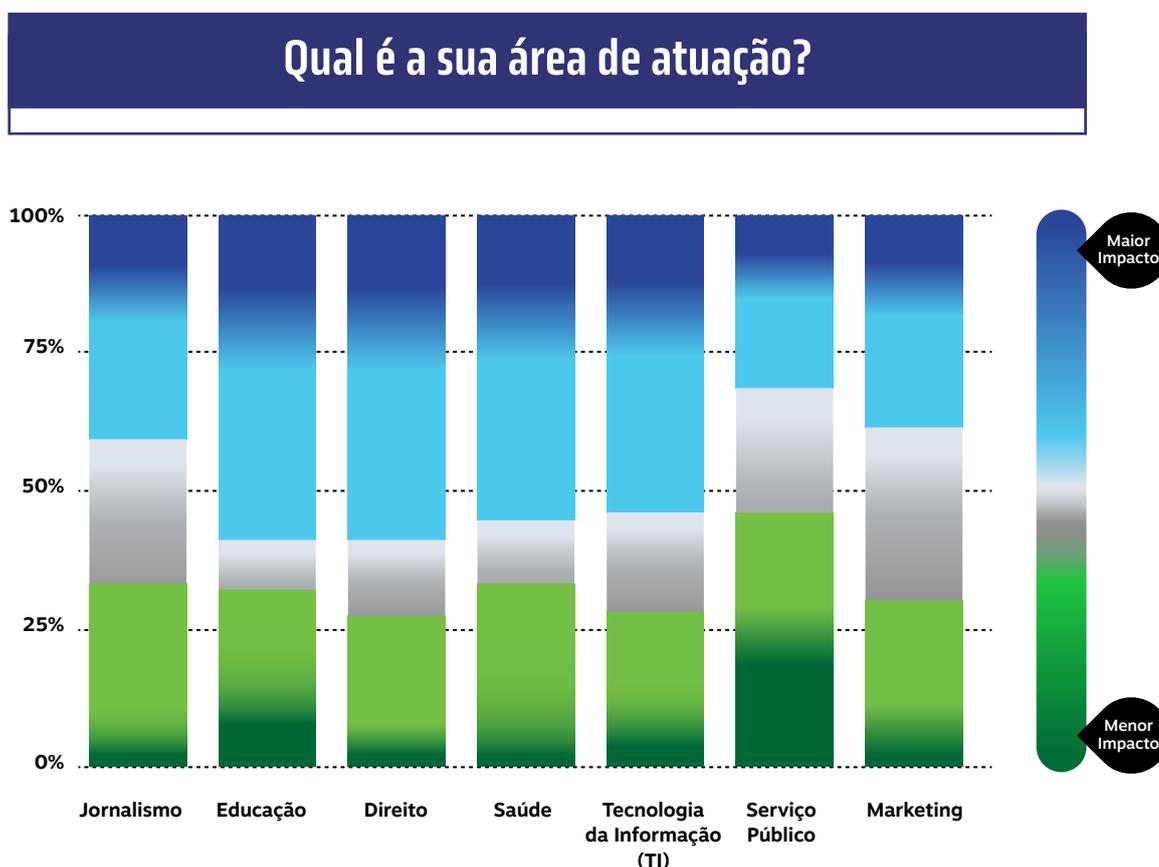
Os textos, sons e as imagens disponíveis para treinamento estão, inclusive, acabando. Estima-se que venham a se esgotar em grande parte nos próximos dois ou três anos. Usando de licença poética, a IA não tem alma ou, sendo mais técnico, ela não passa de um processo estatístico magnífico, mas sem uma compreensão mais profunda do mundo. Como disse recentemente um dos mais renomados cientistas da área, atual chefe do departamento de IA da Meta, Yann LeCun: “Um gato consegue lembrar, consegue entender o mundo físico, planejar ações complexas, pode realizar algum nível de raciocínio — na verdade, muito melhor do que os maiores modelos de linguagem”.

Para navegar nesse oceano do espaço latente da I.A., ainda é preciso uma pessoa, uma intenção, um destino, uma consciência. E deverá ser assim por um bom tempo. Em última análise, a inteligência artificial não é propriamente inteligente.

Ela é, sim, uma excelente ferramenta, talvez a mais poderosa que inventamos até aqui. Uma forma nova e brilhante de organizar, navegar e recombinar o conhecimento humano, seja em forma de texto, imagem, áudio, ou vídeo. Uma ferramenta que inaugura uma nova era nas atividades humanas, que certamente vai reestruturar muitos dos processos de trabalho e criação. Mas ela não vai conseguir fazer isso sozinha, e difere muito de outras inovações anteriores na área de tecnologia. Basta você conversar com ela.

*Este texto foi escrito por um humano, sem nenhuma participação de inteligências artificiais.

Nelson Porto é editor de Inteligência Artificial do Meio e autor das imagens que ilustram esta edição





Heinar Maracy

Qual a ferramenta certa para você?

Existe um mundo de aplicações de IA no mercado. Este guia traz as melhores em quatro categorias: chatbots, criação de imagem, vídeo e áudio

A primeira coisa que você precisa entender antes de começar a utilizar ferramentas de inteligência artificial generativa é que, apesar do nome, elas são burras. Não sabem desenhar, pintar ou escrever criativamente. Apenas reconhecem padrões que identificaram em milhões ou bilhões de textos e imagens e tentam entregar o que você pediu da melhor maneira possível (leia o artigo de Nelson Porto neste relatório para entender melhor como a IA funciona).

A segunda coisa é que essa é uma categoria de software em sua infância, e os aplicativos estão sendo aperfeiçoados enquanto são utilizados. Eles ganham upgrades e novas funções em um ritmo frenético, mas ainda têm

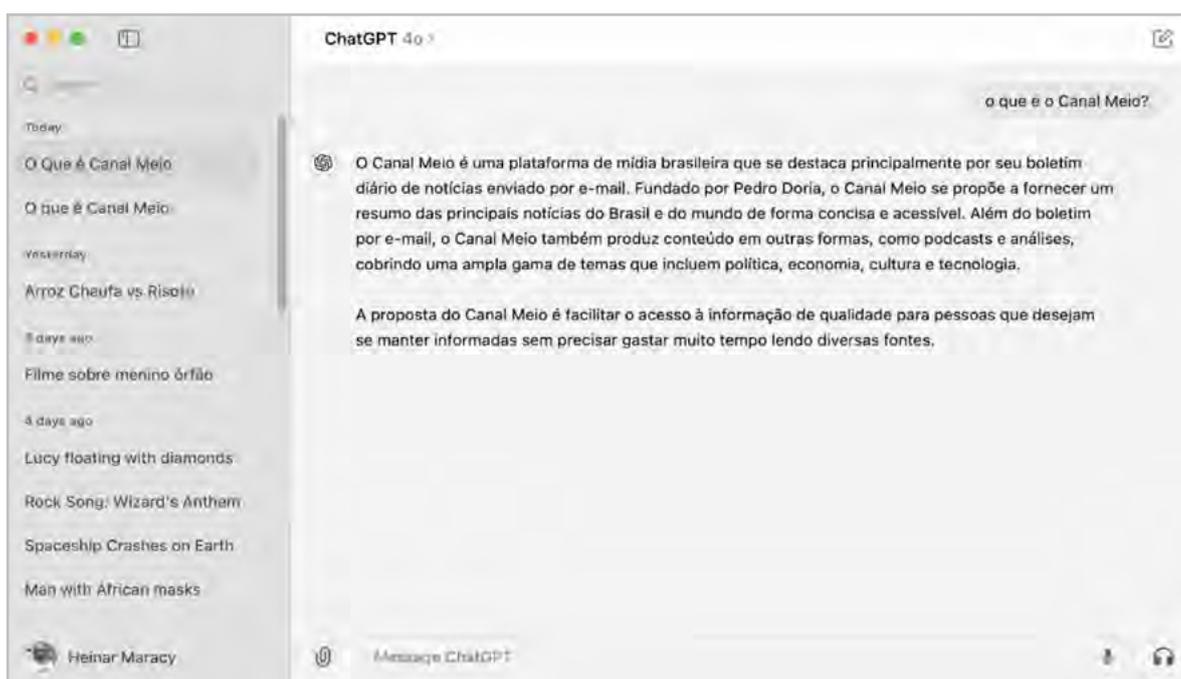
sérios problemas de usabilidade e confiabilidade. São carros sendo construídos enquanto andam e cabe a você decidir se vale a pena dirigir um veículo sem banco ou com cordinhas amarradas nas rodas em vez de volante.

Feita essa ressalva, *aqui estão os volantes mais birutas do mundo para realizar mais uma corrida de IA, numa disputa pelo título de volante mais biruta do mundo*. Dividimos essa disputa em quatro seções: chatbots, criação de imagem, de vídeo e de áudio. Existe um mundo de aplicações de IA generativa no mercado. Vamos nos ater a essas categorias que são as mais populares, segundo nossa pesquisa.

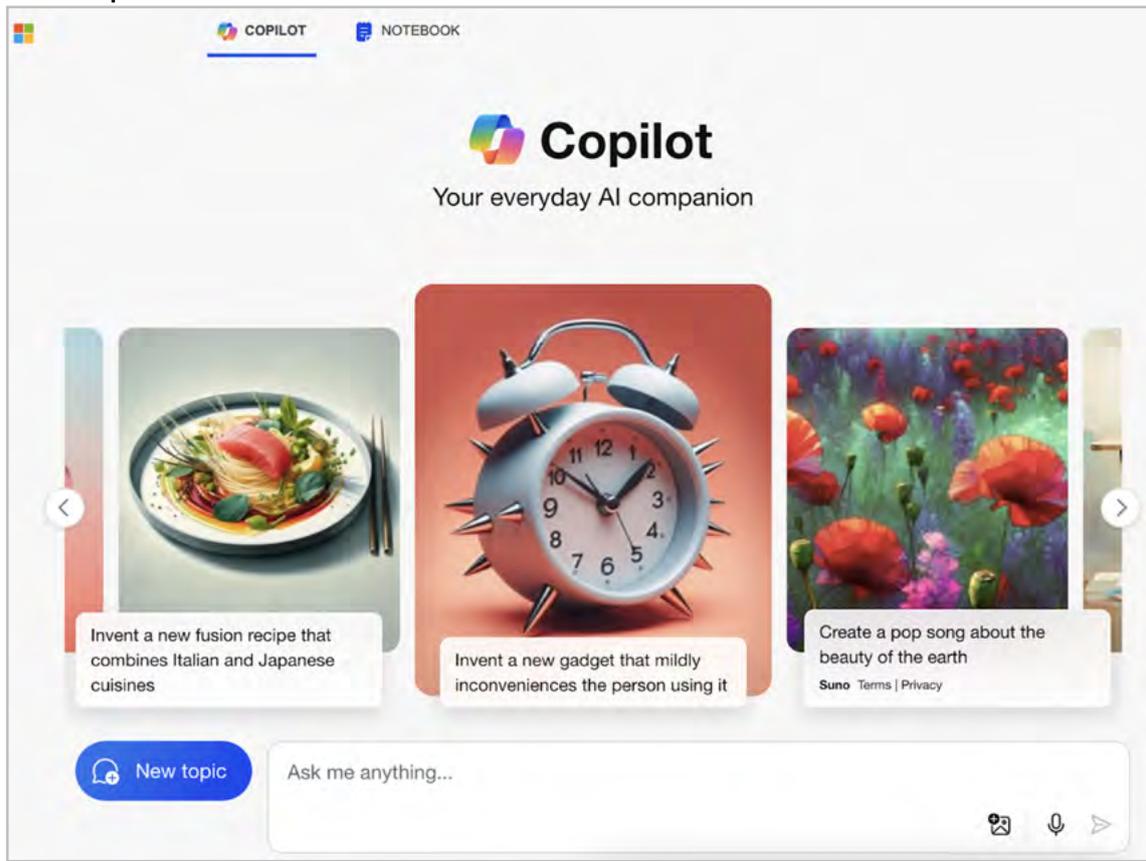
CHATBOTS

Não confunda com aqueles robôs que atendem ligações de cancelamento de serviço ou aparecem no cantinho em páginas web de empresas fingindo melhorar o atendimento. Chatbot na era da IA são interfaces homem-máquina, aplicativos com os quais você conversa em linguagem natural, seja em texto ou voz. O **ChatGPT** é o líder inconteste, com cerca de 70% do mercado de ferramentas de IA. Está tão na liderança que se dá ao luxo de concorrer com ele mesmo. O **Microsoft Copilot** tem em seu motor a mesma LLM do software da OpenAI. Em breve, a Siri da Apple também será conectada a ele, o que deve consolidar sua pole position.

ChatGPT



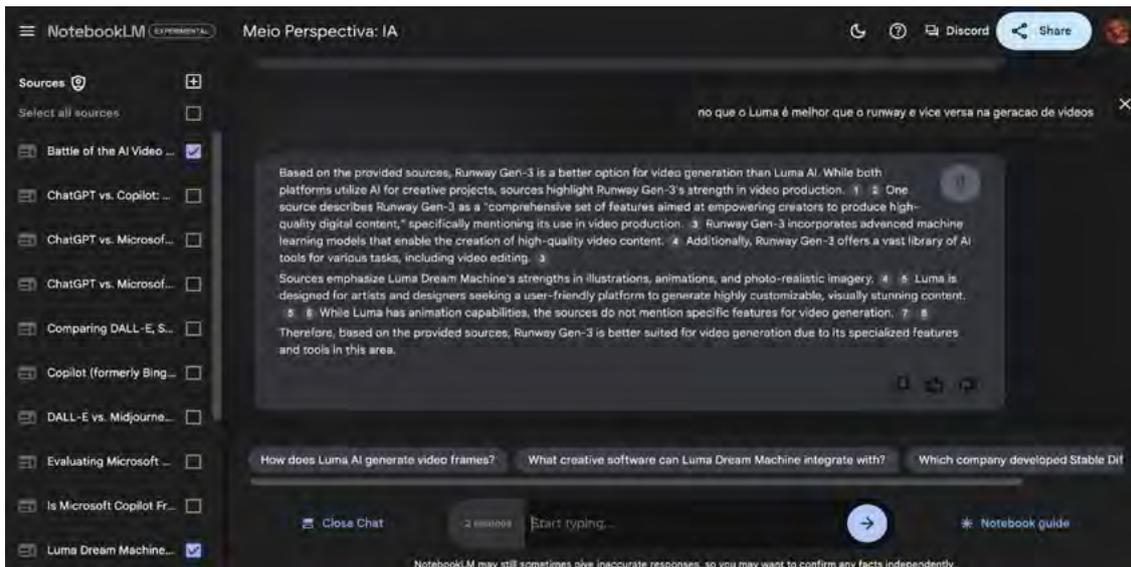
Copilot



Nessa corrida, sair na frente é uma grande vantagem, pois quanto mais gente utiliza o ChatGPT, mais o seu modelo de linguagem é aprimorado. Não se sente confortável sabendo que suas perguntas serão utilizadas para treinar o ChatGPT? Então, use o modo pago ou entre em Configurações > Controlar Dados e desative a opção “melhorar modelo para todo mundo”. A grande vantagem do ChatGPT em relação ao Copilot da Microsoft e ao Gemini do Google é sua flexibilidade e a generosidade do seu modo gratuito, que não tem grandes restrições de uso. A OpenAI vem se esforçando para transformar o ChatGPT em uma plataforma, com aplicativos baseados nele, só que ainda não achou uma *killer application* para chamar de sua.

Nessa corrida de IA, a Microsoft se superou naquilo que faz pior: dar nome aos seus produtos. Pegou o Bing Chat, que era sua versão do GPT, transformou em um assistente virtual, como era a Cortana, e rebatizou de Copilot, antigo nome do assistente de programação do Github. A precificação do Copilot também é uma coisa um tanto confusa. Você pode usar algumas funções dele gratuitamente, pagar US\$ 20/mês para utilizá-lo no

NotebookLM



Office e retirar limites, ou usar a versão Pro (US\$ 360/ano por usuário), direcionada a empresas, embutida no Microsoft 365. Se você utiliza o pacote da Microsoft intensamente em seu fluxo de trabalho, pode ser interessante pagar o Copilot. O ganho de produtividade na elaboração de textos e planilhas é brutal. É bom lembrar que você encontra a mesma funcionalidade com o ChatGPT e seus plugins.

Da mesma forma, se seu *workflow* é baseado nas ferramentas do Google, pode valer a pena utilizar o **Gemini** (antigo Bard). O Gemini não traz grandes vantagens em relação ao ChatGPT (a interface é idêntica, por sinal). Claro, isso pode mudar de um dia para o outro, pois o Google Labs é conhecido por sua rapidez em colocar ideias inovadoras na mão do público. Uma outra ferramenta que merece ser destacada é o **NotebookLM**, ideal para quem escreve textos baseados em múltiplas fontes. Você pode subir PDFs, arquivos de texto ou URLs e pedir para a IA editar, corrigir, sumarizar ou responder perguntas baseadas neles. Uma mão na roda para pesquisadores ou jornalistas. Usei bastante para escrever este artigo.

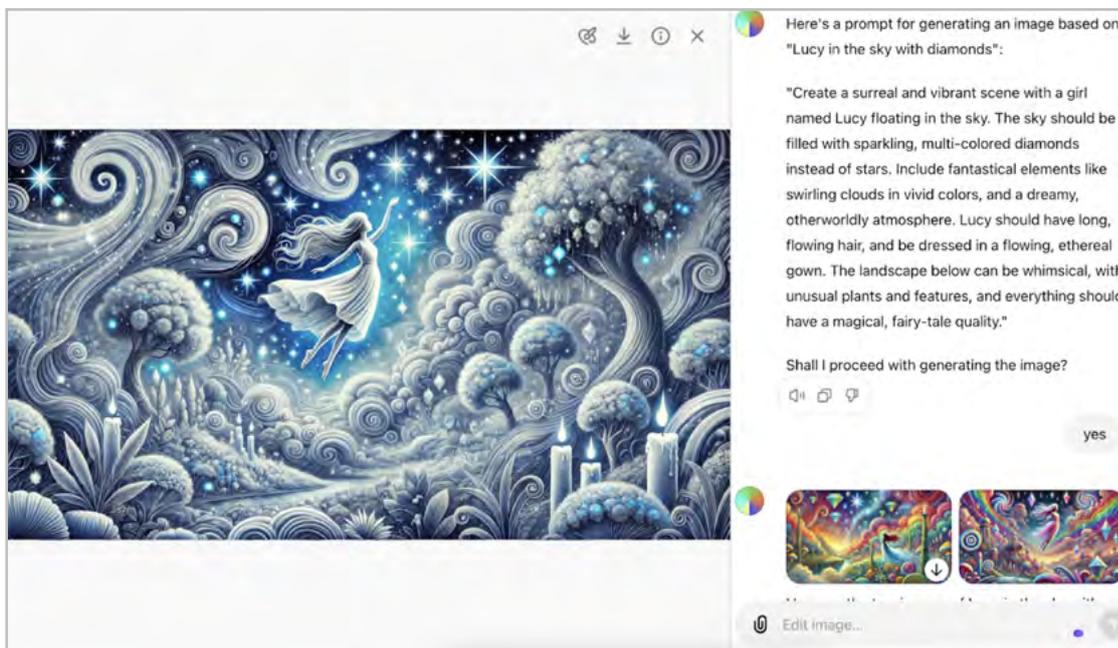
Por último, temos uma ferramenta que ainda não está disponível no Brasil e vale a pena ficar de olho. O **Claude**, da Anthropic, empresa formada por ex-integrantes da OpenAI, só pode ser utilizado por aqui com o auxílio de uma VPN. Aparentemente, seu estilo de escrita é mais humano e menos genérico que o do app da OpenAI e é menos propenso a “alucinações”. Quem já o utilizou afirma que é um dos melhores chatbots para programação.

CRIAÇÃO DE IMAGEM

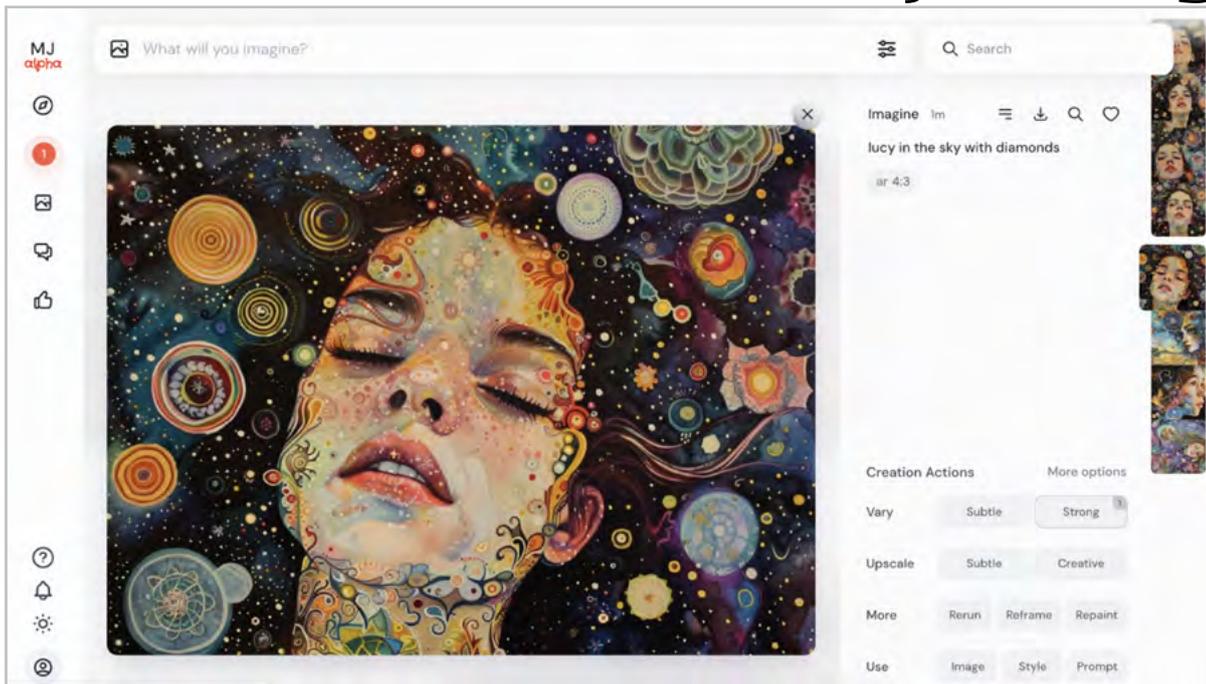
Ferramentas de criação de imagens formam a categoria mais polêmica da nova geração de aplicativos inteligentes. São treinados com imagens de origem oculta e permitem criar “arte” (bota aspas nisso) a partir de comandos de texto. Ou seja, você não precisa mais ter o talento de um Romero Britto para ser um artista plástico ou um ilustrador de sucesso. A comunidade artística ficou irada com quem usa essas ferramentas com o temor de serem ainda mais precarizados, no entanto, é muito difícil que uma AI venha a substituir o talento humano. O mais provável é que ela dê mais condições aos artistas produzirem mais e melhor.

Toda tecnologia revolucionária é ridícula quando nasce. As ferramentas de criação de imagem com IA renderam muitos memes com sua incapacidade de contar corretamente dedos e membros humanos. Apesar de o estigma ter ficado, isso já foi razoavelmente superado. Se por um lado basta uma frase para gerar uma imagem, por outro é praticamente impossível gerar exatamente a imagem que você precisa. Isso não é bug, é feature. Como já foi dito, a IA cria imagens a partir de padrões identificados em uma base de milhões de imagens, decidindo matematicamente qual parte delas se ajusta melhor ao prompt. Por esse motivo, a criação com IA é um trabalho às vezes exaustivo de tentativa e erro. A regra é a seguinte: quanto mais simples a ferramenta, menor o controle que você terá sobre o resultado do seu prompt. Vamos aos principais programas.

Dall-E



Midjourney

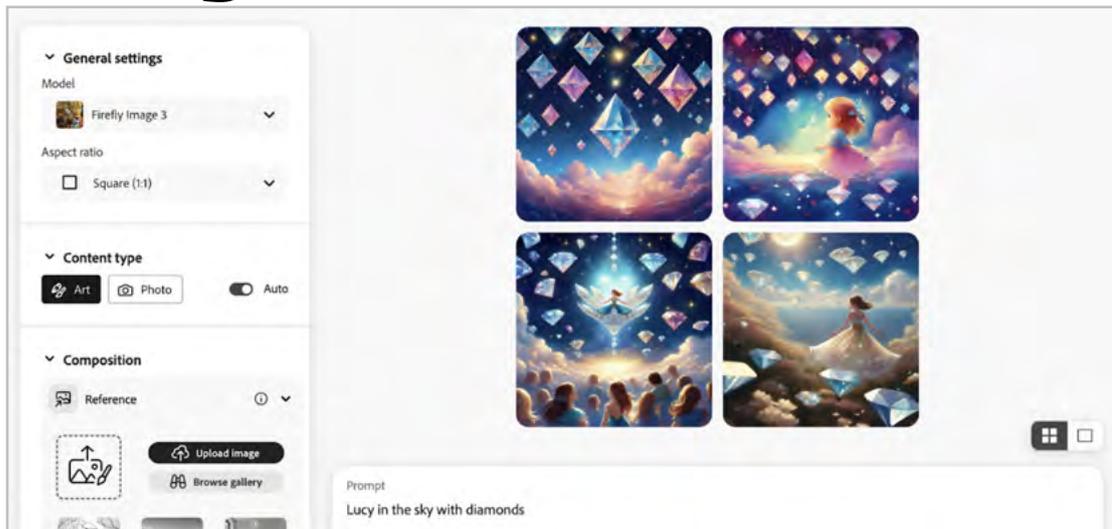


O primeiro aplicativo de geração de imagens a partir de texto foi o **Dall-E**, o irmão artista do ChatGPT, que também é o motor do Copilot Designer. A interface é a mesma: um campo de texto. Sua última versão, o Dall-E 3, só está disponível para quem paga o ChatGPT Plus, mas você pode utilizá-la gratuitamente (com limites diários) no Copilot.

ComfyUI



Firefly

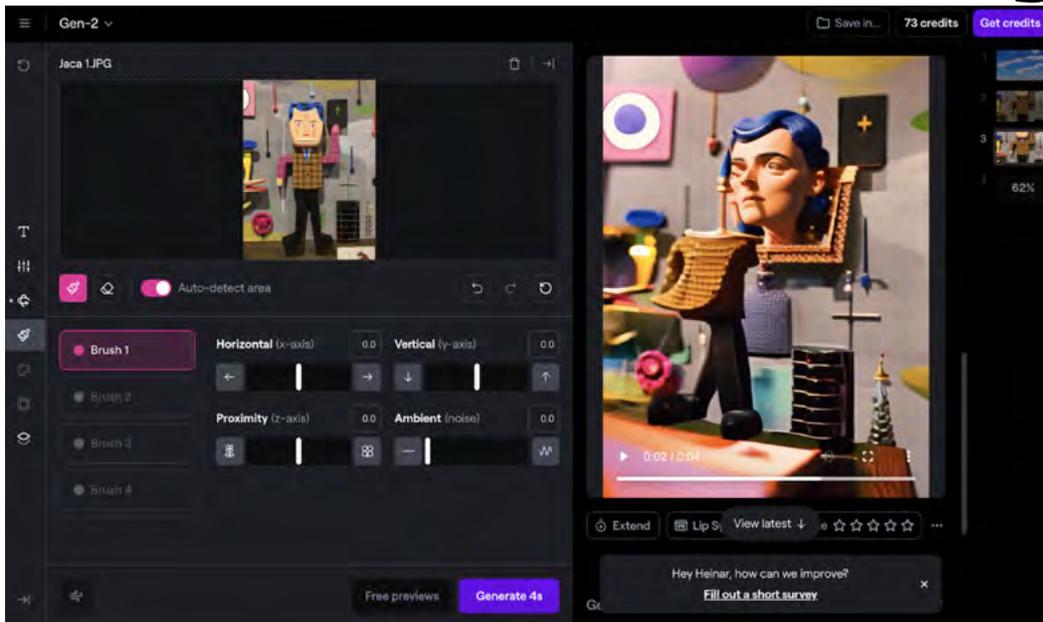


Bem mais flexível e com mais opções que o Dall-E, o **Midjourney** traz, por outro lado, uma curva de aprendizado um pouco maior. Há poucos meses, ele era muito mais difícil de usar, obrigando você a entrar em um grupo de Discord – um misto de programa de mensagens e rede social que só quem tem menos de 25 anos sabe usar – e aprender comandos no braço. Hoje ele tem uma versão alfa que pode ser acessada via browser, muito mais intuitiva. É o aplicativo mais balanceado em termos de flexibilidade e controle dos resultados e facilidade de uso.

Stable Diffusion é o programa que separa as profissionais que trabalham com IA em seu dia-a-dia das garotas que só querem se divertir. É o que oferece mais ferramentas para você domar as incertezas ao criar uma imagem, só que isso requer galgar uma curva de aprendizado íngreme, principalmente se você precisa de consistência entre várias imagens. Ao contrário dos concorrentes, ele é open source e pode ser rodado localmente se você estiver disposto a gastar com seu PC o equivalente a um carro popular. Não tem versão gratuita para experimentar e boa parte de sua funcionalidade é feita por linha de comando ou por uma interface chamada **ComfyUI**, que de confortável não tem nada.

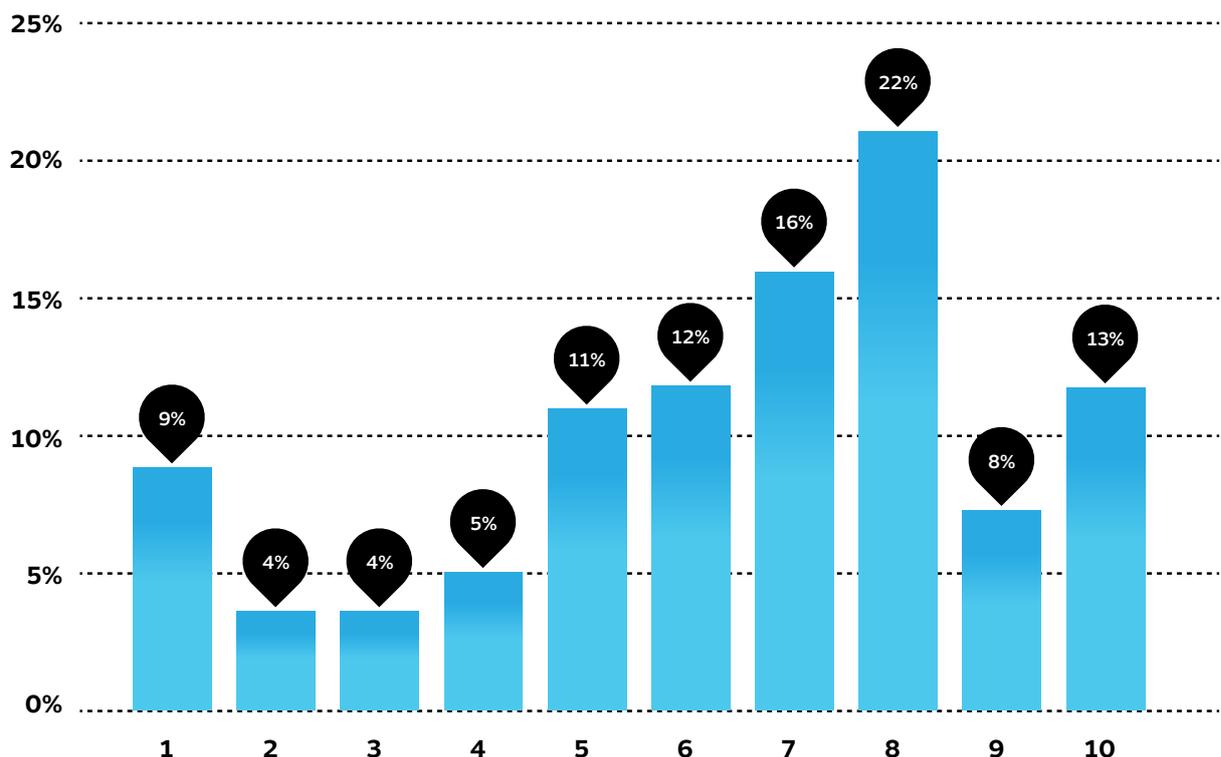
Correndo por fora, temos o **Firefly**, da Adobe. Ele não traz resultados tão espetaculares, talvez por ter sido treinado com um número mais limitado de imagens, as de domínio público e as da Adobe Stock, evitando que artistas que usam os programas da empresa a acusem de estar roubando e plagiando seu próprio público. Totalmente integrado ao Photoshop e Illustrator, é uma mão na roda na produção de protótipos e variações de um projeto.

Runway



Prototipagem, aliás, é a grande aplicação de aplicativos de geração de imagem com IA. No estágio em que se encontram as ferramentas, até é possível criar produtos finais, mas se não houver um conhecimento técnico muito grande ou muitos ajustes no braço, eles ficam com aquela cara de IA, plástica e genérica, que ganhou o apelido de “arte de tiozão”. Como já foi dito, os aplicativos avançam em uma velocidade frenética e isso pode mudar no ano, mês ou semana que vem.

Como você quantificaria o ganho de eficiência gerado pelo uso de IA?



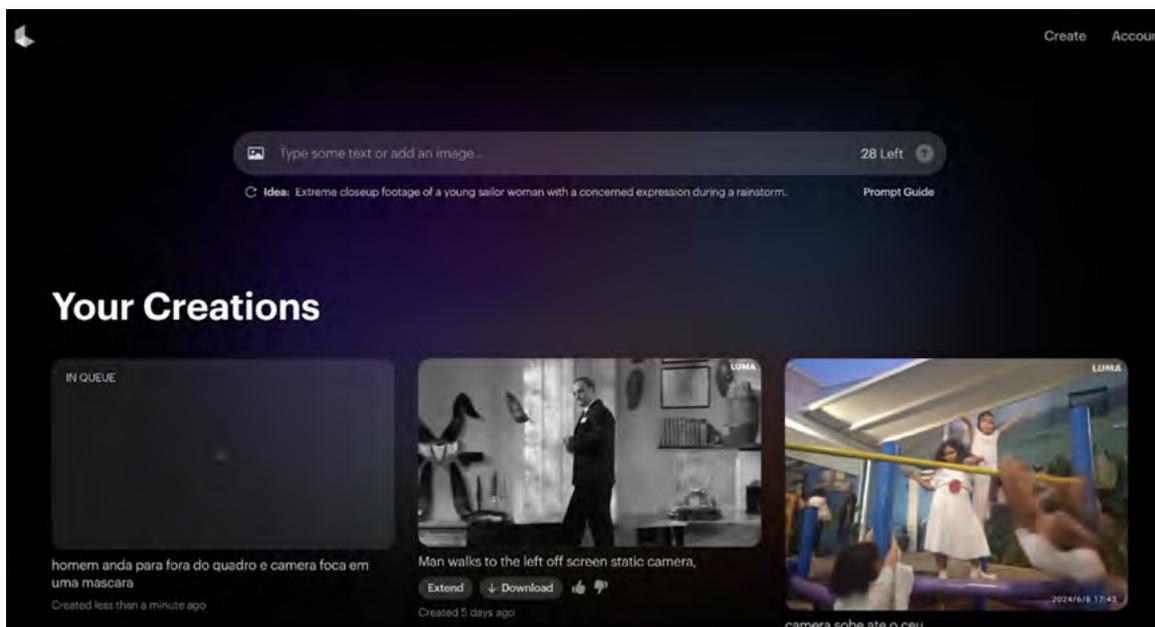
CRIAÇÃO DE VÍDEO

Se controlar o resultado final em uma imagem estática criada por IA é difícil, imagine em 24 imagens por segundo. Os demos de aplicativos de vídeo generativos são fantásticos, a realidade, nem tanto. Para conseguir um resultado específico é necessário um razoável conhecimento de suas entranhas. Não basta dominar “a arte de escrever um bom prompt”. É preciso conhecer a fundo ferramentas e módulos cuja interface está longe de ser considerada amigável.

Enquanto escrevia este artigo, quem estava na frente na corrida do vídeo gerado por IA era o **Runway**, pois seus desenvolvedores colocaram nas mãos do público, prematuramente, uma versão alfa do software, a Gen-3, com medo de que a concorrência roubasse seus clientes. O Runway Gen-3 impressiona na qualidade dos vídeos gerados a partir de texto, mas ainda não tem a possibilidade de criar um vídeo a partir de uma determinada imagem (Image-to-Video) que existe na versão anterior. O Gen-2 é atualmente a ferramenta mais flexível para geração de vídeos, contando com um motion brush para você definir exatamente que parte de uma imagem ele deve animar, lip-sync com vozes pré-programadas ou uma criada por você mesmo e um monte de ferramentas para modificar e criar imagens.

Seu principal concorrente atual é o **Luma**, equivalente em termos de qualidade e aderência ao prompt, só que com menos recursos e flexibilidade. O Luma tem uma

Luma



interface espartana: um campo onde você pode escrever seu prompt ou adicionar uma imagem e transformá-la em vídeo. Não requer prática, tampouco experiência, o que falta é a flexibilidade do Runway.

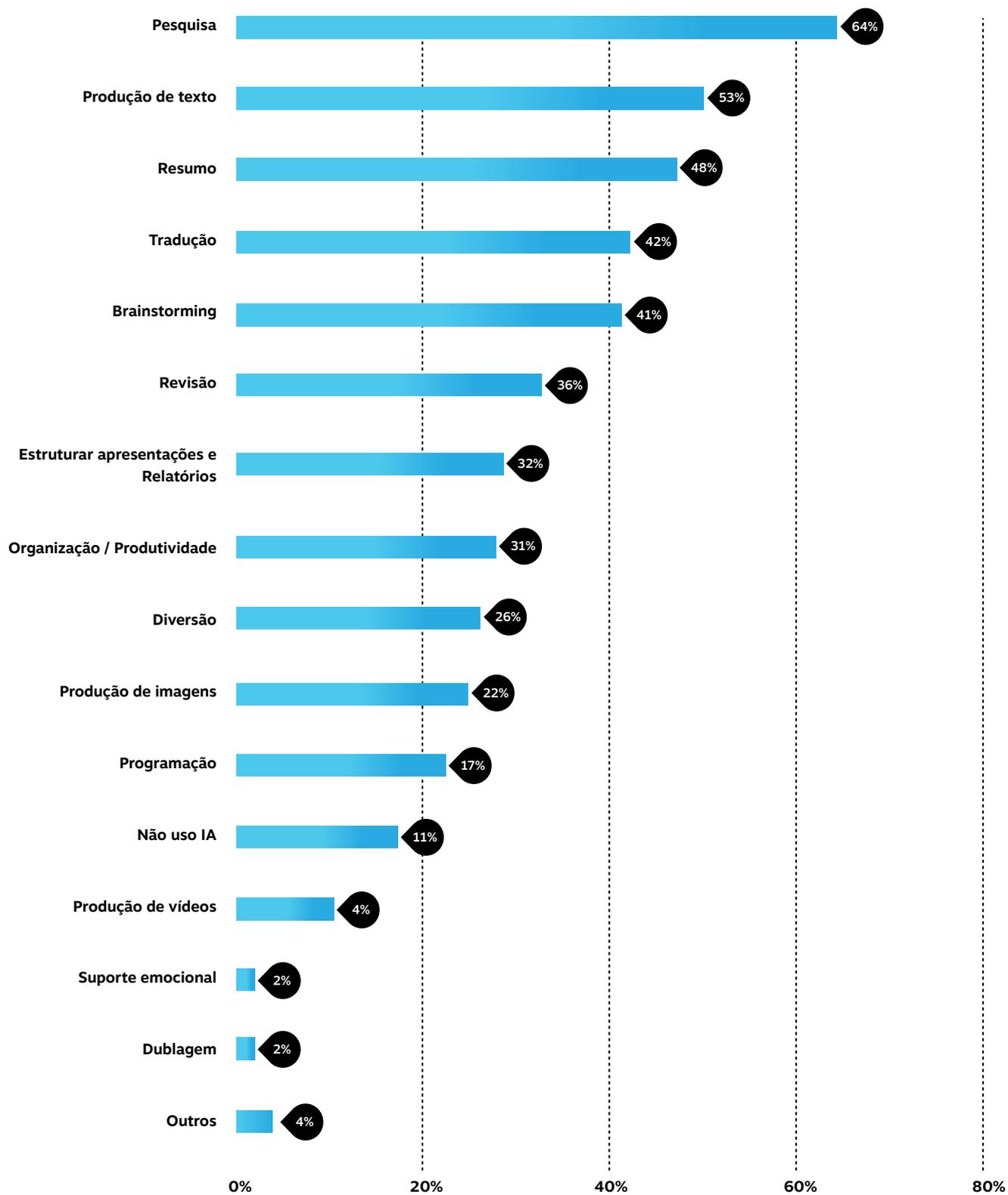
Outro que corre por fora é o **Kling**, que tem gerado exemplos fantásticos. Até o momento, só possui versão em chinês. Entusiastas ocidentais destemidos já conseguiram utilizá-lo baixando um app de celular chinês para validar uma conta em seu site com a ajuda do tradutor do Google. Melhor esperar um pouco. Outro app aguardado ansiosamente é o **Sora**, da OpenAI. Seu anúncio foi o que deu a largada da atual corrida do vídeo generativo e está previsto para ser lançado ainda este ano.

Existem dezenas de projetos sendo lançados em busca do graal do vídeo generativo: a possibilidade de produzir filmes com qualidade suficiente para passar na TV ou mesmo no cinema, o que ainda está um pouco distante (ano que vem, talvez). O grande salto qualitativo será quando ferramentas tradicionais de edição de vídeo incorporarem a IA em seus fluxos de trabalho. Tarefas repetitivas ou que exigem muita proficiência, como motion tracking e tratamento de cor, vão ficar muito mais fáceis com um robzinho ajudando. Provavelmente a próxima edição do Adobe Creative Suite trará plug-ins para Runway, Sora e outros apps generativos.

Kling



Como você usa Inteligência Artificial?



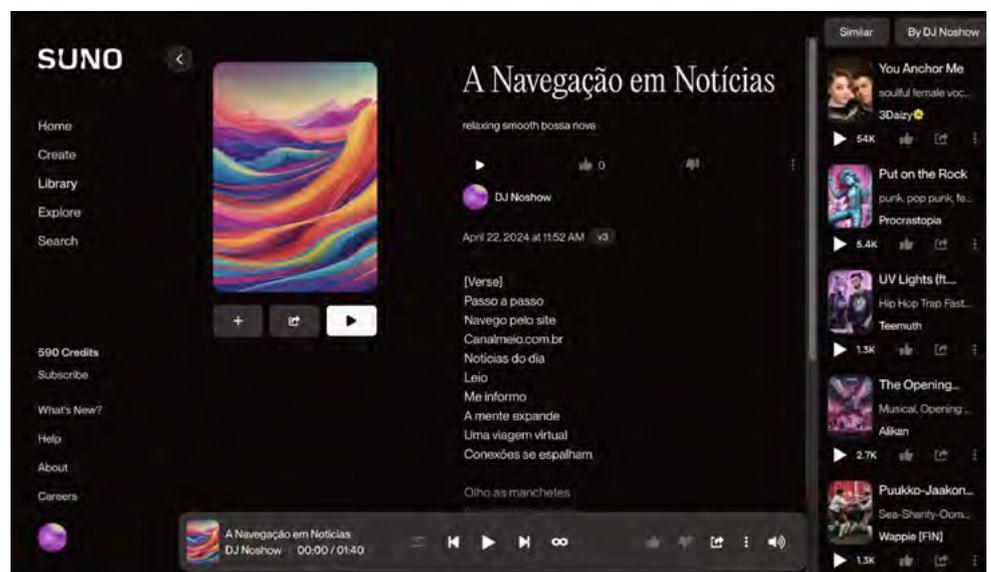
CRIAÇÃO DE ÁUDIO

Como é bom poder tocar um instrumento, dizia Caetano Veloso. Os puristas que reclamavam de DJs e produtores que faziam música apenas apertando botõezinhos em seus computadores devem estar tirando a calça pelo pescoço com o advento da música generativa. Programas como **Suno** e **Udio** permitem que você escreva uma frase e crie uma canção com letra, refrão, instrumentos e backing vocals no estilo e gênero que quiser. Na maioria das vezes, uma música ruim, vamos combinar. Mas é música. E pode ser distribuída pelo Spotify e outras plataformas.

Os programas citados foram alvo de um processo movido por grandes gravadoras, um movimento que lembra os velhos tempos do Napster e do início do MP3. As detentoras dos fonogramas acusam as startups de terem utilizado suas músicas para treinar seus modelos e exigem um pagamento de R\$ 150 mil por música por infração de copyright (o que, vezes alguns milhões, deve dar mais que todo o dinheiro da terra). Os desenvolvedores argumentam que o que a IA faz é ouvir uma música e aprender como seus componentes se relacionam, como faz um músico humano, então estaria fazendo “uso justo” dos tais fonogramas.

Ambos competem corpo-a-corpo para ver quem traz mais funcionalidades. O Suno saiu na frente com o Input Audio, onde você pode subir um arquivo de áudio com você cantando, assobiando ou tocando algum instrumento (como é bom...) para ele criar uma música completa em cima disso. O Udio logo veio atrás, permitindo essa função apenas para usuários pagantes. Quando estava-

Suno





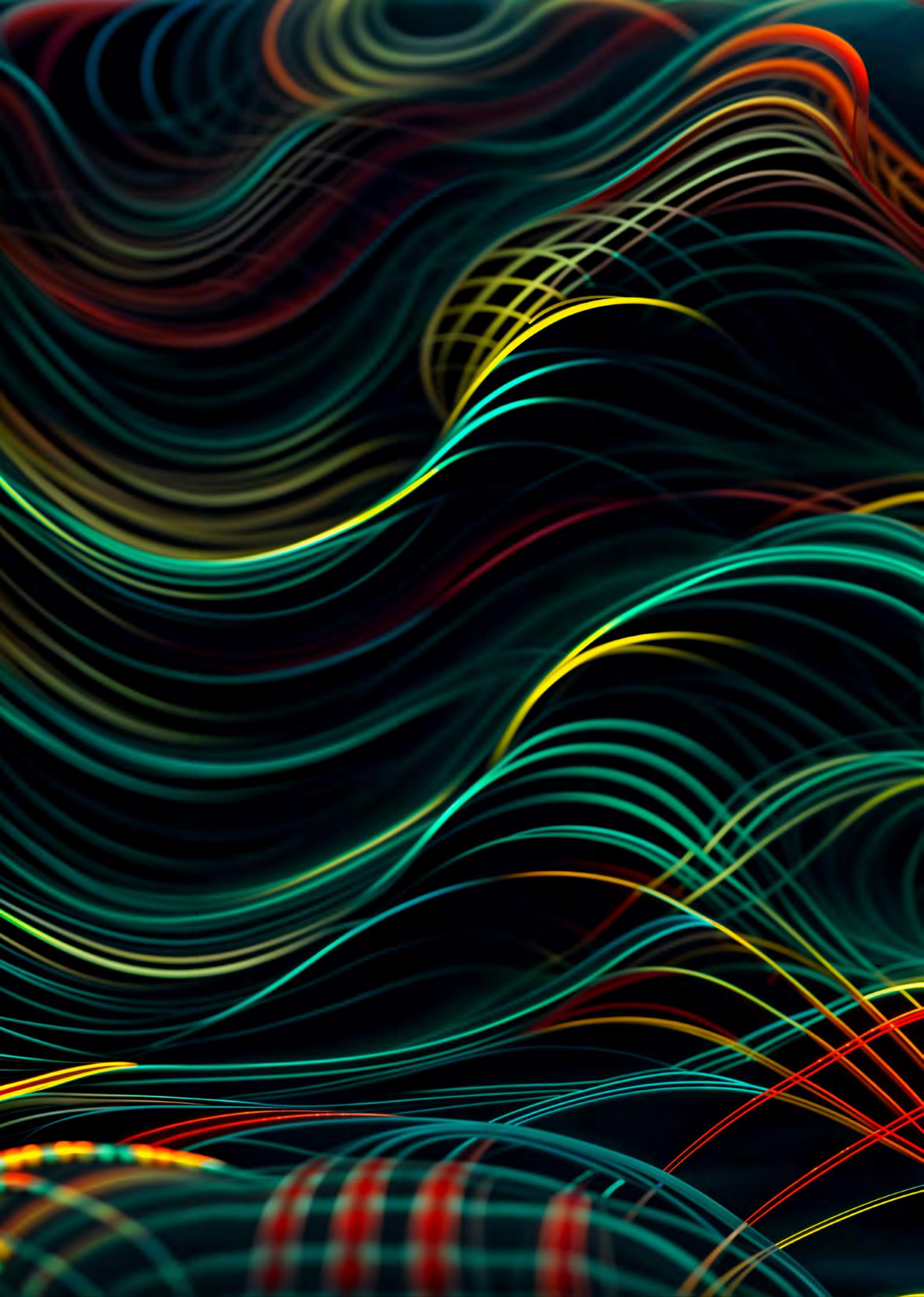
mos escrevendo este artigo, o Suno lançou a possibilidade de gerar músicas em duas pistas independentes e passou à frente novamente, provavelmente por pouco tempo. A qualidade de som do Udio é levemente melhor, mas o Suno se dá melhor com letras em português.

Produtores musicais também irão se beneficiar bastante com a incorporação da IA em seus trabalhos. As ferramentas para manipulação de áudio tornam possíveis coisas impensáveis como recuperar gravações antigas ou mal feitas e separar instrumentos de músicas cujo master se perdeu ou nunca existiu. A Lalal.ia é uma que merece ser conhecida, fazendo gratuitamente essa função, algo que até pouco tempo exigia um software de algumas centenas de dólares.

CONCLUSÃO

Em dois anos, a IA revolucionou a indústria de software. Seus aplicativos trazem possibilidades fantásticas, dando superpoderes a quem precisa programar, escrever, criar imagens, vídeos, planilhas, modelos 3D e impactando praticamente toda atividade que utilize um computador. A maioria sabe fazer só um truque e quase todos ainda têm um longo caminho em termos de usabilidade e controle dos resultados. Mas os aplicativos tradicionais também estão evoluindo, adicionando funcionalidades generativas seja internamente ou com o uso de plug-ins. Não é um absurdo dizer que daqui a um ou dois anos todo software ou plataforma terá algum componente de AI embutido. Pode ser que isso exploda o planeta, mas pelo menos teremos vídeos fantásticos para deixar de recordação.

Heinar Maracy é coordenador de marketing digital do Meio



Guilherme Werneck

A última fronteira

Como está a adoção de inteligência artificial generativa no mundo corporativo

Toda novidade tecnológica passa por uma curva clássica de uso. No topo da pirâmide estão os inovadores, normalmente as pessoas envolvidas na criação de determinada tecnologia. Depois vêm aqueles que a adotam primeiro, que começam a usá-la quando as primeiras versões beta são liberadas para o público. A pirâmide alarga consideravelmente quando a tecnologia atinge as massas, mas ainda assim há dois estágios nesse ponto: o dos que aderem a ela de presto e os que demoram para incluí-la em suas vidas. Por último estão os retardatários, que só a adotam porque já não têm lá muita opção.

Depois de dois anos da disponibilidade das principais aplicações, a inteligência artificial generativa só agora começa a atingir o primeiro estágio de uso em massas. Isso na esfera privada. No mundo corporativo, embora não exista muita dúvida de que essa é uma tecnologia que veio para mudar o mundo como conhecemos, a adoção da IA ainda convive com cenários muito distintos.

Na pesquisa realizada pelo *Meio* junto a seus assinantes para este relatório, quando olhamos para a adoção de IA no ambiente de trabalho de empresas privadas, vemos que ela já é uma realidade incontornável, sem ser muito governada. A vasta maioria, 77%, está numa faixa em que o uso é permitido, mas não é ainda coordenado pelas empresas. O uso estruturado representa 15% e a adoção como parte integral da organização, 7%. Só 3% relatam ter o uso proibido no trabalho. Ainda assim, mais da metade dos entrevistados dizem que já veem retorno financeiro e ganho de eficiência ao adotar a IA no ambiente empresarial.

AMBIENTE CORPORATIVO

Esses dados dialogam com duas pesquisas recentes divulgadas pela IBM neste ano. Uma primeira, feita com profissionais de TI entrevistados em cinco países da América Latina — Argentina, Brasil, Chile, Colômbia, México e Peru — mostra que as corporações da região estão muito abertas ao uso de IA generativa. Dentre as empresas pesquisadas, 37% dizem estar implementando ativamente a tecnologia e 45% afirmam já estar explorando.

O segundo levantamento, feito com CEOs brasileiros, mostra como as lideranças empresariais estão lidando com o assunto. A pesquisa ouviu 3 mil CEOs de mais de 30 países e 26 indústrias diferentes. No que se refere à realidade brasileira, 37% dos CEOs daqui planejam contratar pessoal adicional por causa da IA generativa, 35% pensam que a sua força de trabalho precisará de requalificação nos próximos 3 anos e 82% acreditam que transparência da IA generativa é um elemento chave para conquistar a confiança de clientes.

O diretor de IA, Dados e Automação da IBM Brasil, Thiago Viola, enxerga uma distinção clara no uso de inteligência artificial nas empresas. Ele entende que a grande tendência hoje, no uso de pessoas físicas, é a inteligência artificial generativa, como o Gemini, do Google, e o ChatGPT da OpenAI. Porém, acredita que esse tipo de inteligência artificial não é ainda o coração do uso de IA para negócios, que pedem um ambiente mais controlado. “Quando a gente entra em um espaço de IA for business, tem de

“ Pesquisa da IBM:
37% dos CEOs
brasileiros planejam
contratar pessoal
adicional por causa
da IA generativa ”

tomar alguns cuidados relevantes. Por exemplo, eu não consigo usar um modelo aberto na internet para subir o meu arquivo de fluxo de pagamento da empresa, porque esse dado cai na rede, entra num modelo aberto para todo mundo e eu gero uma inflexão de segurança, de ética, de transparência”, argumenta.

Na IBM, é permitido o uso de IA generativa nas máquinas dos funcionários. Ela tem, inclusive, um modelo proprietário, o Granite, usado em projetos com clientes. Mas, quando pensa no uso corporativo de IA, Viola destaca que é necessário seguir três preceitos fundamentais: ética, governança e transparência. Por isso, hoje a maior parte dos projetos usam inteligência artificial, mas não a generativa, que garanta esses três aspectos, além da proteção dos dados. “Governança é um ponto primordial para que se consiga controlar o espectro inteiro da IA. E aí os módulos de transparência e de ética são subsequentes. Você não pode ter um modelo que não seja transparente. Eu pergunto, como é que vocês chegam a essa conclusão aqui se fulano merece ou não um empréstimo? Preciso explicar, levantar o capô do carro e falar, não, está aqui, cheguei a essa conclusão pelo motivo A,B e C.”

PROJETOS ESTRUTURADOS

Uma empresa que tem trabalhado com projetos estruturados de IA generativa é a Vitru Educação, líder em ensino digital no país, com mais de 900 mil alunos. Como uma companhia que tem a tecnologia em seu DNA, ela já usava diversas ferramentas de inteligência artificial desde 2018, mas passou a olhar mais de perto para as IAs generativas na virada de 2022 para 2023, quando decidiu criar um programa, batizado de Geração, para acelerar o uso interno.

O programa tem três pilares: estratégia e governança, aceleração e capacitação de pessoas e, por fim, experimentação. Foi criada uma série de conversas e comitês para que essa aceleração acontecesse na prática, envolvendo o conselho administrativo, o corpo executivo e um grupo tático. Esse trabalho também promoveu uma aproximação com big techs, como Google e Microsoft, em busca de soluções, além da criação de um programa de inovação aberta para que pudesse, junto a startups, dar mais velocidade aos experimentos.

Algumas dessas experimentações já ganharam o mundo real. Entre elas, um entrevistador virtual, que prepara

os alunos para entrevistas de trabalho, uma ferramenta usada pelo marketing para gerar leads mais qualificados e uma assistente virtual pedagógica, batizada de Sofia, treinada exclusivamente com dados da empresa, já disponível para todos os alunos. Neste momento, a Vitru está também fazendo testes do uso do Copilot, da Microsoft, só que ainda com um grupo seleta de usuários.

“A gente tem patrocínio e encorajamento. Vai testando e estimulando, e esse estímulo segue cascata abaixo”, diz Alessandra Lima, diretora de Inovação da Vitru. “Somos um gigante com mindset de empreendedor. Porque somos uma empresa brasileira, fundada por dois empreendedores, com um DNA de estímulo à execução. Isso ajuda a gente a pensar o que eu posso fazer melhor, mais rápido, mais simples.”

AVENTURA DA IA

Conversamos também com duas lideranças que responderam à pesquisa do *Meio* e estimulam o uso de IA nas suas respectivas empresas, mas ainda não têm projetos mais estruturados em curso.

Marcelo Molnar é sócio-diretor da Boxnet, uma empresa dedicada à análises de dados em comunicação. Entusiasta da inteligência artificial, ele a encara como uma tecnologia que beneficia quem já tem uma certa estrada. “A sensação que eu tenho, por incrível que possa parecer, é que redes sociais foram uma coisa inventada para jovens. Já a inteligência artificial é feita para velhos. Você tem que ter mais conteúdo para explorar, precisa ter um repertório maior.”

Ele enxerga que a IA pode ajudar em processos no trabalho, e, no caso específico da sua empresa, tem usado a tecnologia para a análise de conteúdo, treinando IA para classificar a intencionalidade de textos. Mesmo assim, acredita que ter um analista humano bem capacitado acompanhando e dirigindo o processo e corrigindo imprecisões é fundamental.

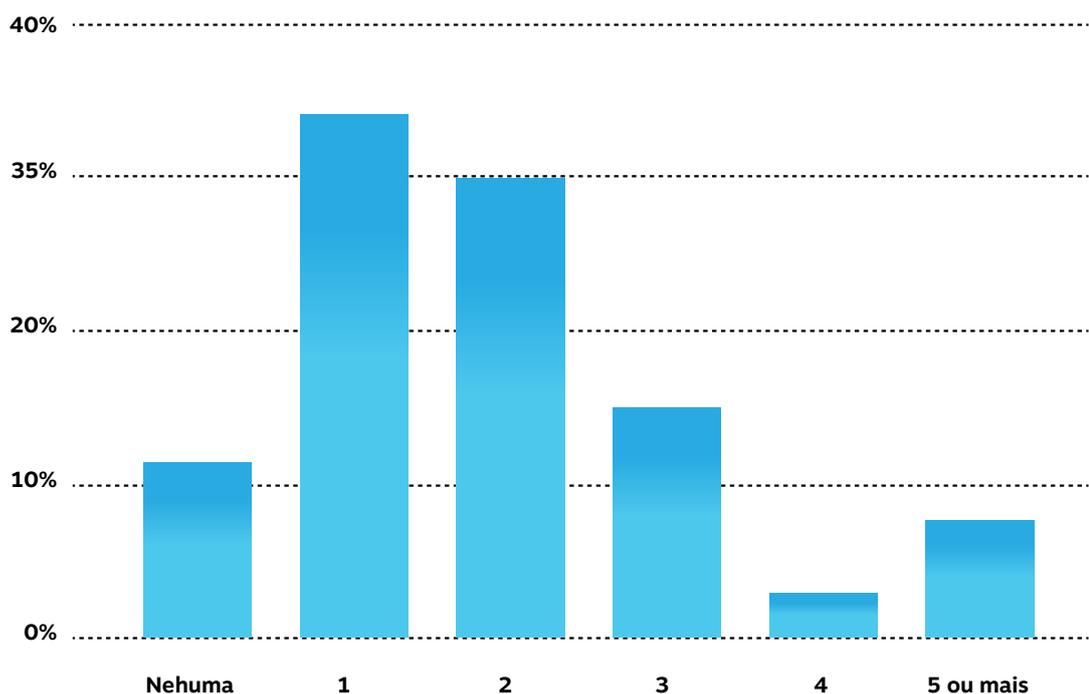
Molnar prevê uma profunda transformação do seu negócio, que hoje conta com cerca de 250 funcionários. Dependendo da área, ele acredita que vá haver mudanças no perfil de contratados. Se alguns tipos de tarefas e funções tendem a desaparecer com o uso da IA, ele acredita que “outras vagas terão de ser criadas justamente por conta desse uso”.

Já o publicitário João Marcelo Meira, sócio e diretor Criativo da agência KOMUH | Garimpo Criativo, de Belo Horizonte, conta que no dia a dia do trabalho o uso de IA é corrente, principalmente nas equipes de criação. “A gente não tem algo estruturado, mas tem um estímulo para a utilização de ferramentas. Liberamos o acesso pago do Midjourney para a equipe, e tem uma utilização muito intensa”, diz. Inclusive, foi criada na agência uma integração entre a ferramenta de comunicação Slack e o Midjourney, com um bot que sugere prompts para usar no aplicativo de criação de imagens. Meira diz ainda que, além de o ChatGPT estar “comendo solto” na equipe, a agência tem pesquisado a produção de vídeo com IA.

Meira também acredita que nos próximos anos veremos uma transformação radical no ambiente de trabalho, defendendo que o primeiro passo vai ser a contratação de pessoas que saibam interagir com IAs. “Eu falo pro meu time o tempo inteiro: quem não abraçar as ferramentas vai ser engolido.”

Guilherme Werneck é editor executivo de conteúdo premium

Quantas ferramentas de IA você usa?





Pedro Doria

A IA super-humana

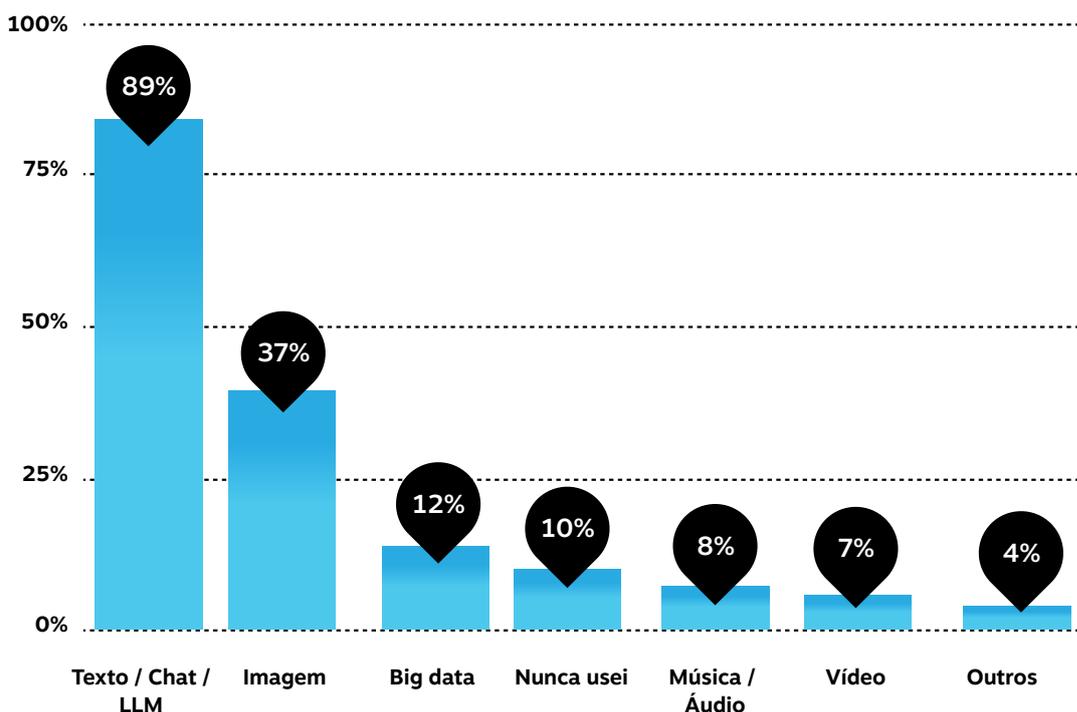
*Quando chegaremos a uma
inteligência artificial que supere
nossos melhores cérebros?*

A Ian Turing, talvez o mais brilhante matemático do século 20, tinha uma ideia de como testar se um computador havia se tornado inteligente. Punha o computador numa sala, uma pessoa na outra, e os dois conversando. Se, após uma sequência de perguntas e respostas, a pessoa não soubesse distinguir se o papo era com gente ou com máquina, havia ali inteligência artificial. O Teste de Turing, como batizamos o exercício, já foi batido inúmeras vezes no arco dos últimos dez anos. É batido todos os dias. Mas aí um fenômeno se deu: embora o teste tenha sido batido, não consideramos ainda que computadores de fato sejam inteligentes. Conforme uma inteligência artificial real se aproxima de ocorrer, vamos nos tornando mais rigorosos, aprimoramos o critério. Então quando é que vai acontecer? Quando é que uma AGI, Inteligência Artificial Geral na sigla em inglês, estará entre nós?

Nos planos futuros da OpenAI, AGI aparece como meta e tem data alvo. É 2027. Estaríamos a três anos dela. Jensen Huang, da Nvidia, sugere algo parecido. Ele fala em uma janela de cinco anos para que ocorra. Essa margem de tempo faz parte de uma crença habitual entre quem trabalha nas empresas envolvidas com o desenvolvimento da tecnologia. Partem do princípio de que, quanto maior a quantidade de dados usados e maior o poder de processamento envolvido no treinamento de um modelo de IA, mais capaz ele se torna. Isto tem sido verdade. Cada geração produz modelos notadamente melhores. Desse jeito, se continuarmos dando mais poder de máquina e mais dados, como vem sendo feito, em algum momento vai acontecer. Ainda assim, tem aí nesse cronograma anunciado por quem é do mercado uma dose de propaganda, de venda de sonhos.

“ Há uma escada por subir, definida em seis degraus por pesquisadores do Google. São: Sem IA, Emergente, Competente, Expert, Virtuoso e Super-humana. Hoje, estaríamos no segundo degrau da escala ”

Quais tipos de IA você usa?



QUANDO VAI ACONTECER?

O Machine Intelligence Research Institute propôs essa pergunta a um grupo de 2.788 pesquisadores da área, todos com papers publicados com a devida revisão por pares em revistas de prestígio. Ao tabular as respostas, que se espalhavam por todo lado, calculou uma chance de 50% de que AGI apareça até 2047. Até 2027, a probabilidade é de 10%.

Ainda assim, definir o que é uma AGI segue não sendo trivial. O que, afinal, quer dizer ser tão ou mais inteligente do que humanos? Quer dizer escrever poesia com a graça de Manuel Bandeira, desenhar como Pablo Picasso, desenhar prédios como Frank Lloyd Wright? Não é o que sugerem os especialistas. A ideia é algo mais próximo de poder compreender os documentos, checar os números e fechar o balanço de uma companhia grande, também gerenciar a agenda de muitos executivos simultaneamente e com precisão, diagnosticar doenças difíceis e, quem sabe, até recomendar livros e séries ao gosto do freguês. Steve Wozniak, fundador da Apple, brinca que para ele já é AGI se a máquina souber entrar na casa, achar a cozinha e fazer um café. (Não é tão trivial assim.) Deste debate no mercado saiu uma definição que vem sendo cada vez mais aceita:

AGI é um sistema capaz de realizar trabalhos variados com real valor econômico, em diferentes áreas, e superando as capacidades de uma pessoa normal na maioria das tarefas.

Se é isto, até lá há uma escada por subir, definida em seis degraus por pesquisadores da DeepMind, empresa de IA do Google. São: Sem IA, Emergente, Competente, Expert, Virtuoso e Super-humana. Hoje, estaríamos no segundo degrau da escala, com inteligência artificial começando a emergir. O passo seguinte, Competente, é uma IA que possa executar as ações tão bem quanto metade dos adultos. O Expert chega mais longe — faz tão bem quanto 90% das pessoas plenamente capazes. O Virtuoso, então, chega ao topo. Melhor do que 99% de nós.

As IAs super-humanas não têm esse nome à toa.

Aprendizados na máquina

Os links das reportagens, pesquisas e dos apps citados neste relatório, para quem quiser se aprofundar

A CORRIDA DA INTELIGÊNCIA ARTIFICIAL

McKinsey: O Estado da IA no início de 2024

MIT: Como a AI generativa vai remodelar as empresas

Stanford: Index IA Relatório Anual 2024

WIPO: Relatório do Cenário Patentes - Inteligência Artificial Generativa

ARTE ARTIFICIAL

Observer: Chefe de IA da Meta, Yann LeCun explica por que um gato doméstico é mais inteligente que a melhor IA

QUAL A FERRAMENTA CERTA PARA VOCÊ?

LLM

Chat GPT

Claude

Copilot

Gemini

Notebook LM

IMAGEM

ComfyUI

Dall-E

Firefly

Midjourney

Stable Diffusion

VÍDEO

Luma

Kling

Runway

Sora

ÁUDIO

Suno

Udio

A ÚLTIMA FRONTEIRA

IBM: 82% dos CEOs brasileiros acreditam que transparência da IA generativa é a chave para conquistar confiança do mercado

IBM: 67% das empresas na América Latina aceleraram a implementação de IA

A IA SUPER-HUMANA

*Machine Intelligence Research Institute :
Milhares de pesquisadores de IA sobre o futuro*

Glossário

Os principais termos que você precisa saber para entender como funciona a inteligência artificial generativa

LLM (LARGE LANGUAGE MODEL)

Os LLMs, ou modelos de linguagem de grande escala, são como superleitores que foram treinados com toneladas de textos. Eles podem prever a próxima palavra em uma frase, ajudando em tarefas como tradução e criação de textos. Esses modelos usam uma estrutura chamada Transformador, que ajuda a entender e gerar texto de maneira coerente, quase como se realmente compreendessem o contexto das conversas.

TOKEN

Tokens são as menores partes do texto que um modelo de linguagem entende — pense neles como sílabas ou palavras quebradas em pedaços menores. A tokenização é o processo de dividir o texto nessas pequenas partes para que o modelo possa processá-lo. A forma como os tokens são usados pode afetar diretamente a eficiência e precisão do modelo.

CONTEXTO

O contexto é toda a informação ao redor de um token que dá sentido a ele. Em modelos como o GPT, manter o contexto é crucial para gerar respostas relevantes. O contexto é limitado pelo número de tokens que o modelo consegue “lembrar” de uma vez. Por exemplo, o GPT-3.5 pode lidar com até 4 mil tokens de contexto, enquanto o GPT-4 pode lidar com até 32 mil tokens. Isso significa que se, durante toda a conversa, você exceder esse limite, o modelo poderá começar a se perder.

RAG (RETRIEVAL-AUGMENTED GENERATION)

RAG é uma técnica poderosa usada quando queremos analisar um volume de dados que ultrapassa o limite de contexto que um modelo de linguagem pode processar de uma vez. Basicamente, quando a quantidade de informações necessárias é maior do que o modelo consegue “lembrar”, o RAG entra em ação. Ele primeiro busca dados relevantes em uma grande base de conhecimento (como se estivesse pesquisando na internet) e depois usa um subgrupo desses dados para gerar respostas informadas. Isso garante que o modelo possa lidar com informações atualizadas e extensas, mesmo que estejam além do seu limite de contexto padrão. É especialmente útil em áreas que precisam de conhecimento constantemente atualizado, como notícias, literatura técnica e pesquisas acadêmicas.

MODELO DE DIFUSÃO

Modelos de difusão são usados principalmente para criar imagens. Eles funcionam pegando uma imagem cheia de “ruído” (como estática na TV) e, passo a passo, removendo esse ruído para revelar uma imagem clara. Esses modelos são ótimos para gerar conteúdo visual, editar imagens e até fazer simulações complexas.

ESPAÇO LATENTE

No mundo do aprendizado de máquina, o espaço latente é como uma versão resumida e simplificada dos dados. Imagine pegar uma foto e reduzir seus detalhes a um conjunto de características principais que ainda capturam a essência da imagem. Isso é feito usando técnicas como autoencoders ou redes neurais generativas adversárias (GANs). O espaço latente mantém as informações mais importantes dos dados originais, mas em um formato mais compacto, facilitando a identificação de padrões e relações. Essa técnica permite que os modelos de IA realizem tarefas complexas, como criar novas imagens, traduzir idiomas automaticamente e reconhecer padrões, de forma mais eficiente e com menos recursos computacionais. O espaço latente é fundamental para muitas aplicações avançadas de inteligência artificial, pois ajuda a entender e manipular dados de maneira mais eficaz.



meio

nomeio.com.br/ia