

## Meta e OpenAI lançarão em breve modelos de IA capazes de 'raciocinar'

Essas versões, dizem as empresas, poderão planejar ações e ter memória, realizando tarefas mais próximas do ser humano

OpenAI e a Meta sinalizam nesta semana que estão prestes a lançar as próximas versões de seus grandes modelos de linguagem (LLMs, pela sigla em inglês), ou seja, os sistemas que alimentam aplicativos de inteligência artificial generativa, como o ChatGPT, revelou o jornal britânico Financial Times.

Segundo executivos das duas empresas, esses novos LLMs serão capazes de raciocinar e planejar, etapas essenciais para alcançar a cognição sobre-humana nas máquinas.

De acordo com o FT, a Meta deve lançar seu modelo, chamado de Llama 3, nas próximas semanas. Por sua vez, a OpenAI, apoiada pela

Microsoft, indicou que seu próximo modelo, que vai se chamar GPT-5, será lançado "em breve".

Em entrevista ao jornal britânico, Joelle Pineau, vice-presidente de Pesquisa de IA da Meta, afirmou que a empresa de Mark Zuckerberg está "trabalhando arduamente para descobrir como fazer com que esses modelos não apenas falem, mas raciocinem, planejem, tenham memória".

Já o diretor de operações da OpenAI, Brad Lightcap, disse ao FT que a próxima geração do GPT mostrará progresso na solução de "problemas difíceis", como o raciocínio. E afirmou que o mundo logo deprecia a ver a IA sendo capaz de realizar tarefas mais complexas de forma mais sofisticada.

Segundo Lightcap, os sistemas atuais de IA são bons em pequenas tarefas, mas suas habilidades ainda são "bastante limitadas". Ele acrescentou que a OpenAI teria "mais a dizer em breve" sobre a próxima versão do GPT.

As atualizações da Meta e da OpenAI fazem parte de uma onda de novos LLMs que surgirá este ano. Outras empresas que preparam-se para lançar são Google, Anthropic e Cohere, lembra o FT.

### INCORPORAÇÃO AO WHATSAPP

A medida que as empresas de tecnologia correm para criar uma IA generativa cada vez mais sofisticada — capaz de criar palavras, imagens, códigos e vídeos de qualidade indistinguível da produção humana — o ritmo do progresso se acelera.

O raciocínio e o planejamento são passos importantes em direção ao que os pesquisadores de IA chamam de inteligência artificial geral — que teria cognição de nível humano —, pois permitem que chatbots e assistentes virtuais conclam sequências de tarefas relacionadas e prevejam as consequências de suas ações.

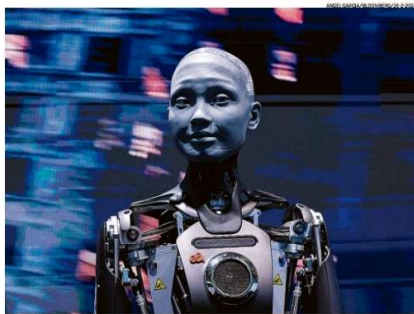
Na terça-feira, em evento realizado em Londres, o cientista-chefe de IA da Meta, Yann LeCun, admitiu que os

sistemas de IA atuais "produzem uma palavra após a outra, realmente sem pensar e planejar". E, por terem dificuldade para lidar com questões complexas ou reter informações por um longo período, ainda acabam cometendo "erros idiotas".

LeCun ressaltou que acrescentar o raciocínio à ferramenta significaria que um modelo de IA "busca respostas possíveis", planeja a sequência de ações e projeta "qual será o efeito" dessas ações. Segundo ele,

essa é uma "grande peça que falta e na qual estamos trabalhando para que as máquinas alcancem o próximo nível de inteligência". Ele citou como exemplo um chatbot de IA que poderia planejar e reservar cada etapa de uma viagem.

A Meta planeja incorporar seu novo modelo de IA ao WhatsApp e aos óculos inteligentes Ray-Ban. A empresa se prepara para lançar o Llama 3 para diferentes aplicativos e dispositivos nos próximos meses.



'Eu, robô': Os novos sistemas de IA serão, cada vez mais, capazes de raciocinar, planejar ações e estimar o efeito delas

## Como 'big techs' usam atalhos para coletar dados

Na corrida para alimentar modelos de inteligência artificial, empresas cortam caminho e ignoram regras de direito autoral

De New York Times  
Sérgio Rosendo

Corrida para liderar a IA tornou-se uma caçada desesperada pelos dados digitais necessários para avançar na tecnologia. Para obter esses dados, empresas como OpenAI, Google e Meta cortam caminhos, ignoram políticas corporativas e debatem sobre contornar a lei.

Na Meta, dona do Facebook e Instagram, gerentes, advogados e engenheiros discutiram no ano passado a compra da editora Simon & Schuster para obter e-mails longos, segundo gravações de reuniões internas obtidas pelo Times. Eles também discutiram a obtenção de dados protegidos por direitos autorais na internet, mesmo que isso significasse enfrentar ações judiciais. Negociar com editoras, artistas, músicos e jornais levaria muito tempo, eles disseram.

Assim como a OpenAI, o Google transcreve vídeos do YouTube para coletar texto para seus modelos de IA, dis-

seram cinco pessoas a par das práticas da empresa. Isso potencialmente violou os direitos autorais dos vídeos, que pertencem a seus criadores.

No ano passado, o Google também ampliou seus termos de serviço. Uma motivação para a mudança, segundo membros da equipe de privacidade da empresa e uma mensagem interna vista pelo Times, foi permitir que o acesso a Google Docs públicos, avaliações de restaurantes no Google Maps e outros materiais on-line para seus produtos de IA.

### BUSCA PELA QUALIDADE

As ações das empresas ilustram como as informações on-line — notícias, literatura, postagens em fóruns, artigos da Wikipedia, programas de computador, podcasts e trechos de filmes — são, cada vez mais, o combustível da próxima indústria de IA. Criar sistemas inovadores demanda ter dados suficientes para ensinar as tecnologias a produzir instantaneamente textos, imagens, sons e vídeos similares

aos criados por um humano.

O volume de dados é crucial. Os principais sistemas de chatbot aprenderam a partir de pools de texto digital abrangendo até 3 trilhões de palavras — o dobro do que existe na Biblioteca da Universidade de Oxford, que armazena textos desde 1602.

Os dados mais valorizados, disseram os pesquisadores de IA, são informações de alta qualidade, como livros e artigos publicados, que foram cuidadosamente escritos e editados por profissionais.

Por anos, a internet, com sites como Wikipedia, era uma fonte aparentemente inesgotável de dados. Mas à medida que a IA avançava, as empresas de tecnologia buscavam mais fontes. Google e Meta, cujos bilhões de usuários pesquisam e postam em redes sociais todos os dias, enfrentavam limites por leis de privacidade e suas próprias políticas, para aproveitar grande parte desse conteúdo.

A situação deles é urgente. As empresas poderiam esgotar os dados de alta qualidade



Pela IA. As empresas de tecnologia usam conteúdo sem negociar com autores

na internet já em 2026, segundo o instituto de pesquisa Epoch. As empresas usam os dados mais rápido do que eles são produzidos.

— O único meio prático para essas ferramentas existirem é se elas puderem ser treinadas em enormes quantidades de dados sem precisar licenciar esses dados — disse Sy Damlé, advogado da empresa de capital de risco Andreessen Horowitz, em um debate sobre direitos autorais no ano passado.

As empresas estão tão ávidas

por novos dados que algumas desenvolvem informações "sintéticas". Não são dados orgânicos criados por humanos, mas textos, imagens e códigos produzidos por IA — ou seja, os sistemas aprendem com o que eles próprios geram.

A OpenAI disse que cada um de seus modelos de IA "tem um conjunto de dados único que selecionamos para ajudar em sua compreensão do mundo". Já o Google afirmou que treina seus modelos de IA "com algum conteúdo do YouTube", o que era

permitido pelos acordos com os criadores do YouTube, que ele controla, e que não usava dados de seus aplicativos corporativos fora de um programa experimental. A Meta, por sua vez, treinou seus modelos com os bilhões de imagens e vídeos publicamente compartilhados no Instagram e no Facebook.

### CÍRCULO VICIOSO

Para os criadores, o aumento do uso de suas obras por empresas de IA tem gerado processos por direitos autorais e licenciamento. O New York Times processou a OpenAI e a Microsoft no ano passado por usar, sem permissão, textos do jornal para treinar chatbots.

As duas empresas argumentaram que esse uso era "justo", ou permitido pela lei de direitos autorais, porque os textos foram usados para outro propósito.

Sam Altman, da OpenAI, já propunha o uso de dados sintéticos desde o ano passado.

Pesquisadores de IA exploram dados sintéticos há anos. Mas construir um sistema de IA que possa se treinar sozinho não é tão fácil como parece. Modelos de IA que aprendem com suas próprias respostas podem ficar presos em um ciclo que reforça suas peculiaridades, erros e limitações.

## Lucro da dona do TikTok dispara 60%, para US\$ 40 bi

Faturamento da ByteDance atinge quase US\$ 120 bilhões em 2023

Do Bloomberg News  
Rafael

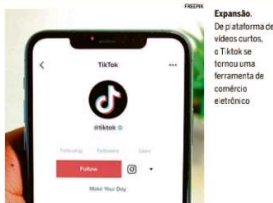
O lucro da chinesa ByteDance saltou cerca de 60% em 2023, superando os ganhos das concorrentes Tencent e Alibaba, em um sinal da resiliência da dona do TikTok, apesar da desaceleração econômica da China e das ameaças de banimento nos Estados Unidos.

O ganho da empresa saltou de US\$ 25 bilhões em 2022 para mais de US\$ 40 bilhões no ano passado, segundo fontes a par do assunto, que não quiseram ser identificadas. A startup mais valiosa do mundo também aumentou seu faturamento de US\$ 80 bilhões para quase US\$ 120 bilhões, disseram as fontes.

Com o TikTok se expandindo dos vídeos curtos pa-

ra o comércio eletrônico, a ByteDance conseguiu, pela primeira vez, ultrapassar a arquirival Tencent tanto em receita quanto em lucro. Embora os números internos da ByteDance não tenham sido auditados de forma independente, eles sugerem que a empresa se tornou um dos gigantes tecnológicos de maior crescimento em 2023.

Um porta-voz da ByteDance



Expansão. De plataforma de vídeos curtos, o TikTok se tornou uma ferramenta de comércio eletrônico

ce não respondeu a um pedido de comentário. Além do TikTok, a empresa tem o Douyin, plataforma de vídeos que só opera na China.

No mercado doméstico, o

Douyin está se consolidando como líder da internet e se transformando em uma plataforma multifuncional semelhante ao WeChat, da Tencent. Seus recursos adicionais

invadem o território de comércio eletrônico da Alibaba e competem com a Meituan na entrega de alimentos. Já o TikTok prevê aumentar em dez vezes o tamanho de seu negócio de e-commerce nos EUA este ano. A plataforma tem 170 milhões de usuários no país.

O maior entrave a isso vem das autoridades americanas. Em março, a Câmara dos Representantes dos EUA aprovou um projeto de lei para proibir o TikTok no país, ameaças que a ByteDance o vende.

A medida, porém, deve enfrentar resistência no Senado. Devido a esse escrutínio, a esperada estreia da ByteDance no mercado de ações ainda é uma possibilidade distante.