

A chinesa DeepSeek ameaça a liderança do ChatGPT?

O mercado de tecnologia foi abalado com a notícia de que um modelo algorítmico desenvolvido pela *Deepseek*, uma companhia chinesa [1], superou o *ChatGPT* [2] em alguns testes de eficiência. O *DeepSeek-R1*, modelo de inteligência artificial generativa [3], atinge desempenho comparável ao *GPT-4 o1*, segundo divulgado [4]. Bateu recorde em número de *downloads*, superando o *ChatGPT* na *App Store* (loja de aplicativos da Apple) e na *Google Play* (da Google) [5].

A notícia ameaçou a confiança na liderança dos EUA no disputadíssimo nicho da inteligência artificial e fez cair as ações das empresas estadunidenses de tecnologia nas bolsas de valores [6]. Em um dia, as empresas norte-americanas de tecnologia perderam US\$ 1,2 trilhão em valor de mercado [7]. Só a *Nvidia* perdeu US\$ 589 bilhões [8].

O que causou todo esse alvoroço foi a divulgação de que o *DeepSeek-R1* foi desenvolvido a um custo cerca 20 vezes menor que os outros modelos das “big techs” americanas, tendo consumido dez vezes menos tempo de treinamento e empregado uma equipe técnica estimada em torno de um terço do número de funcionários das empresas rivais. Divulgou-se que o custo de treinamento do *DeepSeek* foi de apenas 5,58 milhões de dólares, enquanto os modelos das empresas norte-americanas são treinados em regra a um custo entre 40 a 100 milhões de dólares.

Somente grandes empresas de tecnologia, com muitos recursos computacionais e grandes somas de capital para investimento, têm condições de desenvolver e lançar no mercado os chamados *LLMs*. Os grandes modelos de linguagem (*large language models* — *LLMs*), a exemplo do *ChatGPT*, são modelos treinados em grandes quantidades de dados para realizar diversas tarefas, incluindo algumas para as quais não foram especificamente desenvolvidos e treinados [9].

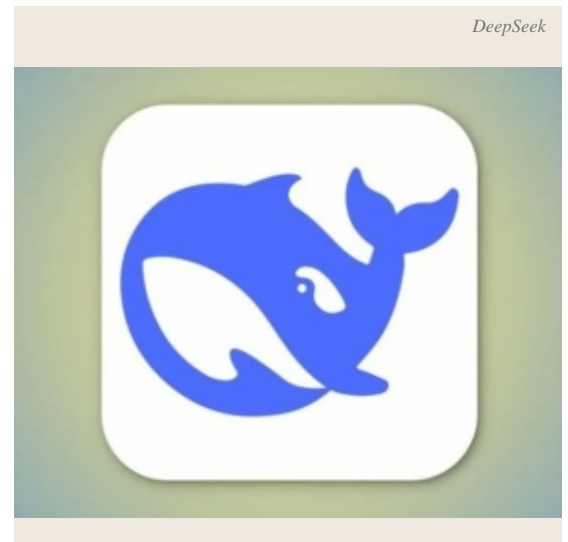
Desenvolver e treinar um modelo de *LLM* requer quantidade substancial de dados, poder computacional, equipe de técnicos e engenheiros de alto nível e, evidentemente, recursos financeiros. A *OpenAI*, em parceria com a *Microsoft*, a *Google* através de sua subsidiária *Deep Mind*, a *Meta* e a *Anthropic* possuem significante poder e recursos para exercer influência nesse ecossistema. Era basicamente esse pequeno número de empresas norte-americanas que vinha desenvolvendo os modelos mais sofisticados de IA Generativa.

Quebra de paradigma

A divulgação da performance do *DeepSeek R1* parece quebrar esse paradigma, mas o mercado de IA dificilmente será aberto para pequenas e médias empresas. O domínio permanecerá nas mãos de um pequeno leque de empresas, as chamadas “big techs”. Não é só desenvolver e colocar no mercado de consumo um modelo fundacional [10] para uma empresa ser alçada ao clube da elite da tecnologia de IA. Especialmente em se tratando de modelos de IA de grande capacidade (*highly capable AI models*), a exemplo dos *LLMs*, o provedor em regra continua com a hospedagem e alimentação do modelo.

O provedor é quem tem recursos para hospedagem em nuvem, centralização dos dados e poder computacional para funcionamento do modelo, atribuições difíceis de serem terceirizadas. O fato novo da divulgada performance do *DeepSeek R1* revela apenas que as empresas norte-americanas podem perder o domínio do mercado de IA para as empresas chinesas ou que, pelo menos, estão tendo sua liderança muito ameaçada. E é nisso que reside a explicação para a queda brutal das ações das big techs norte-americanas.

A boa nova é que possivelmente o acesso a ferramentas de IA fique mais barato. Ultimamente, todas as novas versões de modelos de IA generativa das empresas norte-americanas tiveram seus preços aumentados para o usuário final. A assinatura mensal da versão *Pro* do modelo *ChatGPT o1*, anunciada no início de dezembro do ano passado, já está custando US\$ 200 (cerca de R\$ 1,2 mil) [11]. O modelo *DeepSeek-R1* é de 20 a 50 vezes mais barato de usar [12]. A empresa chinesa tem preços mais baixos em razão do menor custo de produção dos seus modelos. Isso certamente vai pressionar as big techs americanas a reverem suas estratégias de preços.



A política de restrição à importação de tecnologias mais avançadas para desenvolvimento da IA generativa (*IAGen*), implementada ainda durante o governo Biden, parece não ter surtido efeito para a empresa chinesa que criou o novo modelo. Acreditava-se que a vantagem dos EUA no setor de semicondutores e a limitação do acesso à China aos chips mais avançados (como o H100 da Nvidia) garantiria a liderança no desenvolvimento da IA. Mas a empresa chinesa desenvolveu um modelo tão eficiente quanto os das empresas norte-americanas, aparentemente usando chips de menor capacidade.

Novo propósito de IA

O *DeepSeek-R1*, assim como o modelo *GPT* da empresa estadunidense *OpenAI*, a par de ter capacidades generativas, é um exemplo de IA de propósito geral (que vem do termo em inglês *general-purpose artificial intelligence* ou simplesmente *GPAI*). Aliás, os grandes modelos de linguagem (*Large Language Models — LLMs*) enquadram-se nessa acepção da tecnologia *GPAI*. São modelos treinados em grandes quantidades de dados para realizar diversas tarefas [13]. O que a empresa *DeepSeek* fez foi criar métodos mais eficientes para treinar seus modelos, tornando-os mais econômicos ao exigir menos recursos de computação para o treinamento. O modelo mais recente da *DeepSeek* é tão eficiente que exigiu um décimo do poder de computação utilizado para treinar o modelo *Llama 3.1* da *Meta* [14].

O que também pode ameaçar as grandes empresas americanas de tecnologia é o método escolhido pela *Deepseek* para disseminar seu modelo de IA. Ao invés de fornecer uma *API* [15] para que desenvolvedores parceiros construam modelos mais especializados, a *DeepSeek* optou por permitir acesso ao modelo base como um *software* de código aberto (*open source*). Quando decide distribuí-lo como *open source*, o provedor libera a arquitetura para acesso livre ao público, para qualquer um modificar, estudar e usar.

A estratégia de distribuir modelos sob a forma de *open source*, em lugar de restringir o acesso a parceiros escolhidos (por meio de *API*), retira do provedor o controle sobre quem faz uso do modelo. Mas por outro lado pode ganhar em termos de reputação da empresa e participação colaborativa da comunidade de programadores no aperfeiçoamento do modelo. A decisão da *DeepSeek* de lançar seu modelo como *open source* pode democratizar o acesso à tecnologia, promover colaborações e, potencialmente, aumentar a base de usuários.

Ainda é muito cedo para afirmar que o modelo da *DeepSeek* vai tomar mercado dos modelos proprietários das empresas norte-americanas. O modelo da *DeepSeek* é configurado para não responder ou evitar perguntas sobre temas sensíveis ao governo chinês [16]. Isso pode ser um desincentivo para ser usado por pessoas residentes fora da China ou por quaisquer outras que não comunguem com os padrões de censura impostos. Existem também preocupações quanto à segurança dos dados pessoais dos usuários, que poderiam ser indevidamente explorados pelo governo chinês [17].

O que o surgimento do *DeepSeek* deixa claro é que os EUA têm um concorrente de peso na corrida pelo domínio da IA. A China emerge como o principal concorrente dos EUA e o lançamento *DeepSeek R1* pode ser considerado o marco temporal da escalada da “guerra fria tecnológica” entre os dois países.

[1] www.deepseek.com/

[2] O *ChatGPT* é uma ferramenta algorítmica que imita a linguagem natural, um tipo de inteligência artificial conversacional, ou seja, um *chatbot* que conversa e estabelece diálogos com o usuário. Para saber mais sobre o *ChatGPT*, sugerimos a leitura de nosso artigo intitulado “O fenômeno do *ChatGPT* desperta a necessidade da regulamentação da IA”, publicado no site *Conjur*, em 19.03.23, acessível em: <https://www.conjur.com.br/2023-mar-19/democrito-filho-necessidade-regulamentacao-ia/>

[3] A empresa chinesa também disponibiliza o modelo *DeepSeek-V3*, lançado em dezembro de 2024.





[4] Ver artigo divulgado pela *DeepSeek*, onde discorre sobre a performance do seu modelo de raciocínio em comparação com o modelo da *OpenAI*: <https://arxiv.org/pdf/2501.12948>

[5] Ver notícia publicada pela *Blomberg*, em 27.01.25, disponível em: <https://finance.yahoo.com/news/china-deepseek-tops-iphone-downloads-032520148.html>

[6] Ver notícia publicada no *Infomoney*, em 27.01.25, disponível em: <https://www.infomoney.com.br/mercados/ia-chinesa-deepseek-lidera-downloads-no-iphone-e-acoes-de-techs-desabam/>

[7] Ver notícia publicada na *Folha de São Paulo*, em 27.01.25, disponível em: <https://www1.folha.uol.com.br/mercado/2025/01/acoes-de-empresas-de-tecnologia-perdem-us-1-tril-em-valor-de-mercado-apos-anuncio-de-ia-chinesa.shtml>

[8] Ver notícia publicada no *O Globo*, em 27.01.25, disponível em: <https://oglobo.globo.com/economia/tecnologia/noticia/2025/01/27/nao-foi-so-a-nvidia-indice-de-empresas-de-tecnologia-nos-eua-perde-us-1-trilhao-em-um-dia-com-avanco-chines-em-ia.ghtml>

[9] Os modelos de linguagem são usados para compreender e responder a perguntas em línguas naturais, como o inglês, o português, o francês, o espanhol etc. Para gerar textos em língua natural, os modelos de linguagem são treinados para aprender a prever a próxima palavra ou frase com base no contexto anterior. Eles podem ser usados para várias tarefas, como tradução automática, geração de texto, resumo automático e resposta a perguntas. Os modelos de linguagem mais recentes, como o *GPT-4* da *OpenAI*, são baseados em redes neurais profundas e apresentam um desempenho impressionante em várias tarefas de processamento de linguagem.

[10] Para melhor entendimento sobre como funcionam os modelos fundacionais, sugerimos a leitura de nosso artigo “IA de propósito geral e modelos fundacionais: dificuldades para regulação”, publicado no site Conjur, em 17.06.24, disponível em: <https://www.conjur.com.br/2024-jun-17/inteligencia-artificial-de-proposito-geral-e-modelos-fundacionais-as-dificuldades-para-regulacao-dessas-novas-tecnologias/>

[11] O novo serviço pago é dez vezes mais caro que o anterior, *ChatGPT Plus*, que continuará sendo oferecido por US\$ 20. Ver notícia publicada no *TechTudo*, em 05.12.24, acessível em: <https://www.techtudo.com.br/noticias/2024/12/openai-lanca-chatgpt-pro-por-r-12-mil-por-mes-saiba-mais-edsoftwares.ghtml>

[12] O modelo disponibilizado gratuitamente pela empresa chinesa é o *DeepSeek-V3*.

[13] Devido a essa característica dos grandes modelos de linguagem, de se prestarem a realizar variadas funções, podem ser utilizados como aplicação de modelo-base para outros sistemas de IA. As tecnologias de IA com capacidades generativas podem ser empregadas como estruturas de base para outros sistemas, com o mínimo de ajuste fino.

[14] De acordo com pesquisa realizada pela *Epoch AI*, divulgada no seguinte link: <https://epoch.ai/gradient-updates/how-has-deepseek-improved-the-transformer-architecture>

[15] API significa *Application Programming Interface* (Interface de Programação de Aplicação). A interface pode ser entendida como uma ponte ou elo de ligação entre duas aplicações ou sistemas informáticos diferentes.

[16] Segundo testes realizados por usuários, o *DeepSeek* desvia de perguntas sobre a Praça da Paz Celestial, sobre o Presidente Xi Jinping ou a possibilidade da China invadir Taiwan.

[17] Ver notícia publicada no jornal inglês *The Guardian*, em 28.01.25, disponível em: <https://www.theguardian.com/technology/2025/jan/28/experts-urge-caution-over-use-of-chinese-ai-deepseek>

Fonte: <https://conjur.jumps.com.br/2025-fev-03/a-chinesa-deepseek-ameaca-a-lideranca-do-chatgpt/>