Editora: Ana Paula Macedo
anapaula.df@dabr.com.br
3214-1195 • 3214-1172

Por uma linguagem mais humana

Cientistas norte-americanos e espanhóis pesquisam meios para que dispositivos de inteligência artificial desempenhem a chamada generalizada composicional, habilidade própria das pessoas de ampliação de aprendizado e da comunicação

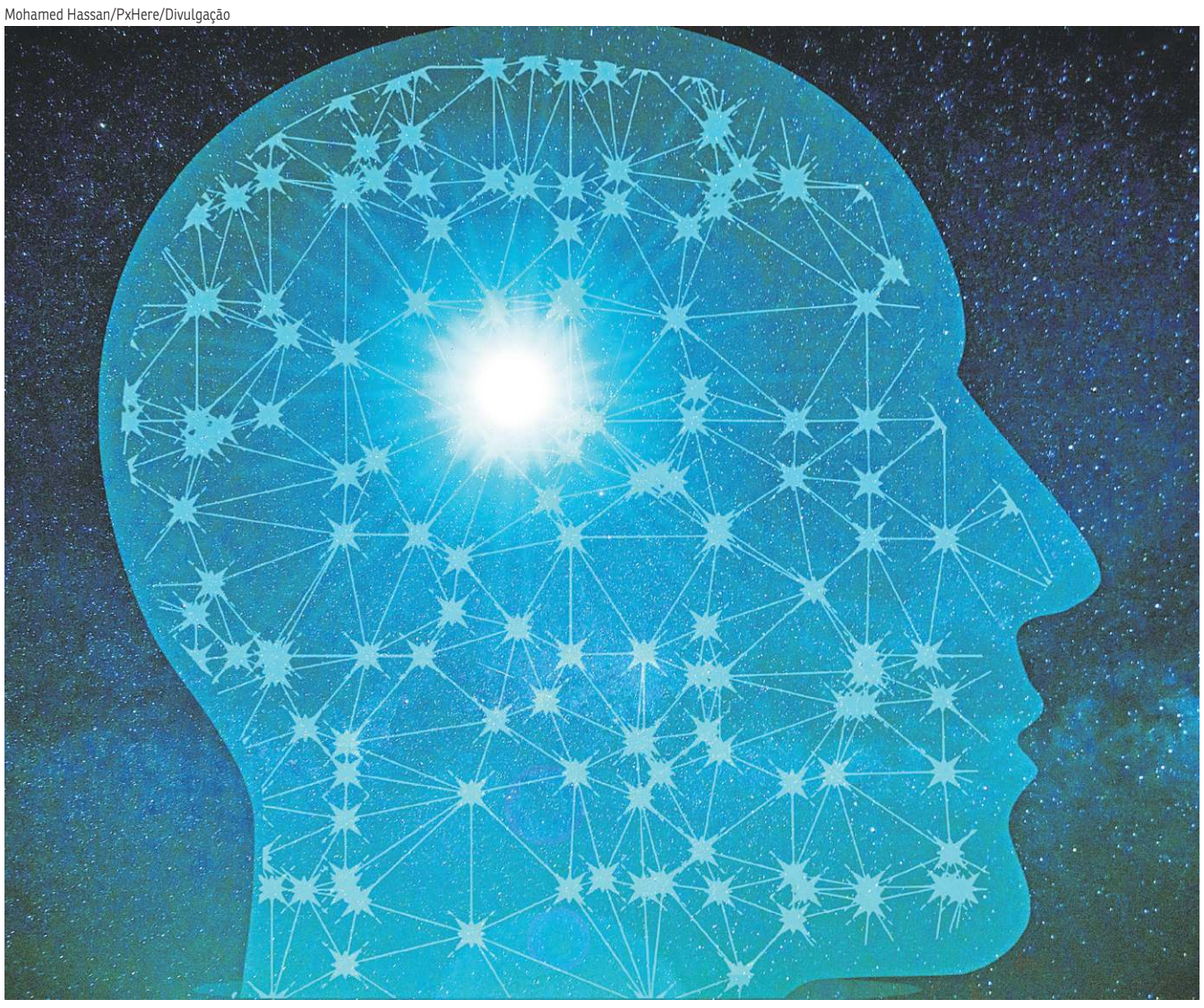
» PALOMA OLIVETO

Os humanos têm uma capacidade inata de compreender novos conceitos e, uma vez adquiridos, utilizá-los em contextos diferentes. Por exemplo, se uma criança é ensinada a andar nas pontas dos pés, ela entenderá imediatamente o que é correr nas pontas dos pés. A habilidade, chamada de generalização composicional, era atribuída unicamente ao *Homo sapiens*. Até que o "homo máquina" também começa a dominá-la.

Recentemente, pesquisadores espanhóis e norte-americanos relataram, na revista *Nature*, uma técnica pioneira, com potencial para desenvolver a generalização composicional em sistemas computacionais no mesmo nível ou até melhor que as habilidades humanas. Segundo os autores, isso não significa que as máquinas ficarão mais inteligentes do que seus criadores. Brenden Lake, da Universidade de Nova York, nos Estados Unidos, e um dos autores, explica que o objetivo é melhorar as ferramentas generativas de inteligência artificial, como o já popular ChatGPT, incrementando a interação com humanos.

A publicação do artigo ocorre 40 anos depois das primeiras tentativas de ensinar a generalização composicional às máquinas. No fim da década de 1980, os filósofos e cientistas cognitivos Jerry Fodor e Zenon Pylyshyn propuseram que essa seria uma habilidade impossível para a inteligência artificial. Então, diversos cientistas começaram a tentar provar que eles estavam errados e passaram a desenvolver redes neurais artificiais com esse propósito. Até agora, Fodor e Pylyshyn ganhavam o debate.

Mohamed Hassan/PxHere/Divulgação



Um sistema tecnológico que busca evitar as ferramentas degenerativas da inteligência artificial e que visa a comunicação

Para saber mais

Aprendizado profundo

As redes neurais artificiais (RNAs), ou redes neurais simuladas (SNNs), são um subconjunto do aprendizado de máquina e estão no centro dos algoritmos de deep learning. Seu nome e estrutura são inspirados no cérebro humano, imitando como os neurônios biológicos se comunicam.

As RNAs são compostas por uma camada de nós: uma, uma ou mais ocultas e uma de saída. Cada nó, ou neurônio artificial, se conecta a outro e tem um peso e um limite associados. Elas dependem de dados de treinamento para aprender e melhorar sua precisão ao longo do tempo. No entanto, uma vez que esses algoritmos são ajustados, tornam-se ferramentas poderosas, permitindo classificar e agrupar informações em alta velocidade.

As tarefas de reconhecimento de fala ou de imagem podem levar minutos ou horas quando comparadas à identificação manual por especialistas humanos. Uma das redes neurais mais conhecidas é o algoritmo de busca do Google. (**PO**)

Fonte: IBM

Caminho histórico

1943

Warren S. McCulloch e Walter Pitts publicaram Um cálculo lógico das ideias imanentes na atividade nervosa. *A pesquisa procurou entender como o cérebro humano poderia produzir padrões complexos por meio de células cerebrais conectadas, ou neurônios. Uma das principais ideias que surgiu do trabalho foi a comparação de neurônios com um limite binário com a lógica booleana (ou seja, 0/1 ou declarações verdadeiro/falso).*

1958

Frank Rosenblatt é creditado com o desenvolvimento do perceptron, documentado em sua pesquisa, O Perceptron: Um Modelo Probabilístico para Armazenamento e Organização de Informações no Cérebro. *Ele leva o trabalho de McCulloch e Pitt um passo adiante ao introduzir pesos na equação. Aproveitando um IBM 704, Rosenblatt conseguiu fazer com que um computador aprendesse a distinguir os cartões marcados à esquerda dos daqueles à direita.*

1974

Embora vários pesquisadores tenham contribuído para a ideia da retropropagação, Paul Werbos foi a primeira pessoa a observar sua aplicação em redes neurais em sua tese de doutorado.

1989

Yann LeCun publicou um artigo ilustrando como o uso de restrições na retropropagação e sua integração na arquitetura da rede neural podem ser usados para treinar algoritmos.

Fonte: IBM

Padrão

Os desenvolvedores dos sistemas de aprendizado de linguagem têm insistido em ensinar a generalização composicional às ferramentas com os métodos de treino-padrão, alimentando as máquinas com dados sucessivamente. Outros chegaram a criar novas arquiteturas, para que as máquinas adquirissem a habilidade. Porém, a nova tecnologia, chamada Meta-aprendizagem para Composicionalidade (MLC, sigla em inglês), aposta em exercícios específicos, praticados até o aprendizado. Algo como ocorre com os humanos.

No MLC, a rede neural artificial é atualizada continuamente para melhorar suas habilidades ao longo de diversas fases. Na primeira, a máquina aprende uma nova palavra e deve aplicá-la em uma frase. Por exemplo, a palavra "pular". A ferramenta, então, precisa criar uma sentença como "pular duas vezes". Feito isso, é estimulada a avançar ainda mais, e assim sucessivamente.

"Grandes modelos de linguagem como ChatGPT continuam a ter problemas com generalização composicional, embora tenham melhorado nos últimos anos", afirma Marco Baroni, professor do Departamento de Ciências da Tradução e da Linguagem da Universidade Pompeu Fabra, na Espanha. Ele é o coautor do artigo. "Mas acreditamos que o MLC pode melhorar ainda mais as habilidades composicionais de grandes modelos de linguagem", afirma.

Para testar a eficácia da tecnologia, Lake e Baroni conduziram uma série de experimentos com participantes humanos, nos quais foram apresentadas tarefas idênticas às realizadas pelo sistema MLC. Além disso, em vez de aprenderem o significado de palavras reais — termos que as pessoas já conheceriam —, foi ensinado aos voluntários o significado de termos inexistentes, como zup e dax, cujos conceitos foram desenvolvidos pelos pesquisadores. Os participantes tinham, então, que aplicar os conceitos em contextos diferentes. Baroni diz que o desempenho do MLC foi tão bom — e, em alguns casos, melhor — que o dos humanos.

Dificuldade

Tanto o MLC quanto os voluntários superaram o ChatGPT e o GPT-4, que, apesar de apresentarem habilidades surpreendentes em termos gerais, demonstraram dificuldade na tarefa de aprendizagem ligada à generalização composicional.

"Mostramos, pela primeira vez, que uma rede neural genérica pode imitar ou superar a generalização sistemática humana em uma comparação direta", diz Lake, professor assistente do Centro de Ciência de Dados e do Departamento de Psicologia da Universidade de Nova York. "Essa é uma questão de décadas, nas quais pesquisadores em ciências cognitivas, inteligência artificial, linguística e filosofia têm debatido se as redes neurais podem alcançar uma generalização sistemática semelhante à humana."

Professor do Departamento de Informática da Universidade de Valladolid, na Espanha, Teodoro Calonge avalia o artigo como "engenhoso", mas alerta que ainda é um experimento inicial. "É oportuno esclarecer que este é um primeiro trabalho. Não saberia dizer se é uma linha de pesquisa que vai oferecer avanços no curto ou médio prazo", admite.

Em seguida, Calonge acrescenta que: "É claro que não creio que esse artigo responderá às questões atualmente levantadas sobre o domínio da inteligência artificial". Porém, o especialista acredita que redes neurais artificiais mais habilidosas poderão ajudar especialmente as pesquisas biomédicas, com ganhos para a humanidade.

Ferramenta impede gravações de voz falsas

Há pouco mais de um mês, a plataforma TikTok foi invadida por notícias falsas, acompanhadas por áudios que imitavam, com bastante similaridade, a voz de famosos. Uma vítima foi o ex-presidente Barack Obama. O vídeo mostrava o político norte-americano se defendendo de uma suposta teoria conspiratória sobre a morte de seu ex-chefe de cozinha. Não foi um fato isolado. Com o aprimoramento das ferramentas de inteligência artificial, as clonagens de som têm se tornado cada vez mais comuns.

Para combater fraudes do tipo, pesquisadores da Universidade de Washington, em St. Louis, desenvolveram uma ferramenta, a AntiFake, que impede a síntese de fala não autorizada. A novidade foi apresentada na Conferência sobre Segurança de Computadores e Comunicações em Copenhague, na Dinamarca.

Ning Zhang, professor assistente de ciência da computação e engenharia da instituição norte-americana, destaca que os avanços recentes na inteligência artificial generativa estimularam desenvolvimentos na síntese de fala realista. Embora a tecnologia tenha o potencial de melhorar a acessibilidade de pacientes com prejuízos comunicativos, além de tornar mais agradável os assistentes de voz, ela também levou ao surgimento dos *deepfakes* como as declarações falsas atribuídas a Obama.

Ao contrário dos métodos tradicionais de detecção de deepfake, que são usados para avaliar e descobrir áudio sintético como uma ferramenta de mitigação pós-ataque, o AntiFake tem uma postura proativa. Ele emprega técnicas para impedir a síntese de fala enganosa, tornando mais difícil para as ferramentas de IA ler as características necessárias das gravações de voz. O código está disponível gratuitamente para os usuários.

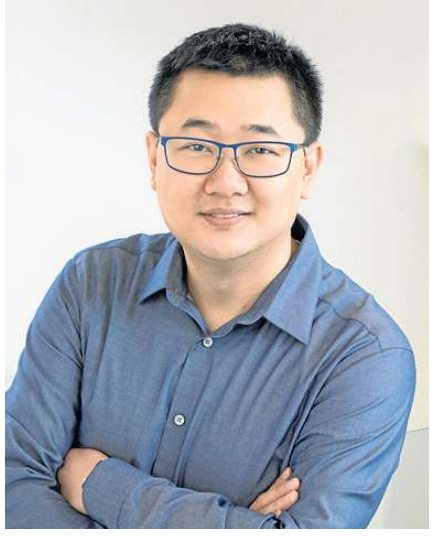
"O AntiFake garante que, quando divulgamos dados de voz, seja difícil para os criminosos usarem essas informações para sintetizar nossas vozes e se passar por nós", disse Zhang, em nota. "A ferramenta usa uma técnica de IA adversária, que originalmente fazia parte da caixa de ferramentas dos cibercriminosos, mas agora a estamos usando para nos defendermos deles", conta. "Nós bagunçamos um pouco o sinal de áudio gravado, distorcemos ou perturbamos apenas o suficiente para que ainda soe bem para os ouvintes humanos, mas é completamente diferente para a IA."

Mudanças

Para garantir que o AntiFake possa resistir a um cenário em constante mudança de invasores e modelos de síntese desconhecidos, Zhang e o primeiro autor do artigo, Zhiyuan Yu, um estudante de pós-graduação no laboratório do pesquisador, construíram a ferramenta para ser generalizável e a testaram contra cinco sintetizadores de fala. O AntiFake alcançou uma taxa de proteção superior a 95%, mesmo contra programas comerciais inéditos. Eles também testaram a ferramenta com 24 participantes humanos, para confirmar que ela é acessível a diversas populações.

Atualmente, o AntiFake pode proteger clipes curtos de fala, visando o tipo mais comum de representação de voz. Mas, segundo Zhang, não há nada que impeça esta ferramenta de ser expandida para proteger gravações mais longas, ou mesmo música, na luta contínua contra a desinformação. "Por fim, queremos proteger totalmente as gravações de voz", disse Zhang. "Embora eu não saiba o que virá a seguir na tecnologia de voz de IA — novas ferramentas e recursos estão sendo desenvolvidos o tempo todo —, acho que nossa estratégia de usar as técnicas dos adversários contra eles continuará a ser eficaz."

Washington University/Divulgação



"A IA generativa estimula a fala realista", diz o pesquisador Ning Zhang