

IA na saúde vai bem na teoria, mas falha com pacientes reais, mostra estudo

- *Pesquisa publicada na Nature Medicine revela que leigos acertam menos diagnósticos ao usar chatbots*
- *Pesquisadores alertam para a importância do letramento e da supervisão médica no uso da tecnologia*

Cláudia Collucci

São Paulo

Embora grandes modelos de IA (inteligência artificial), como GPT-4o, Llama 3 e Command R+, atinjam ótimos desempenhos em exames médicos teóricos, eles não melhoraram a capacidade de leigos de tomar decisões corretas diante de sintomas comuns.

Em alguns casos, o uso das ferramentas até piorou a identificação de condições clínicas importantes, o que pode representar um risco à saúde. O alerta vem de um estudo publicado na revista Nature Medicine em meio ao entusiasmo crescente do uso da IA na saúde.

O trabalho, conduzido por pesquisadores da Universidade de Oxford e colaboradores, testou 1.298 adultos no Reino Unido. Cada participante recebeu um de dez cenários clínicos —situações como dor de cabeça súbita e intensa, falta de ar, dor abdominal, febre persistente ou sintomas gripais.

A partir daí, cada um precisou decidir que tipo de atendimento procurar: autocuidado, consulta de rotina, atendimento urgente, pronto-socorro ou chamar a ambulância.

Os participantes foram divididos em quatro grupos: três puderam conversar com um dos modelos de IA; o quarto, grupo controle, usou os recursos que normalmente utilizaria em casa —sobretudo mecanismos de busca e o site do sistema público de saúde britânico.

Antes de envolver pessoas, os pesquisadores testaram os modelos isoladamente. Sozinhos, eles identificaram ao menos uma condição médica relevante em cerca

de 95% dos casos. Quando solicitados a indicar a conduta adequada, acertaram em média 56% das vezes.

Mas o cenário mudou quando humanos passaram a interagir com as IAs. Eles identificaram corretamente condições relevantes em menos de 34,5% dos casos —índice inferior ao do grupo controle. Na escolha da conduta, não houve diferença estatisticamente significativa: todos acertaram pouco, em torno de 43%.

Para o médico Matheus Ferreira, diretor de IA no Super Professor, uma plataforma educacional, o estudo expõe um ponto central: a diferença ali é o interlocutor. "Quando o médico fez a pergunta, veio quase 100%. Quando o paciente fez a pergunta, veio em 30 e poucos por cento", afirma.

O problema, diz, não está apenas na tecnologia, mas em como ela é usada. "O paciente chega e vai jogar lá: 'tenho dor na barriga, o que pode ser?'. Ele não é um especialista em prompt, não descreve passo a passo."

Como os modelos são probabilísticos, acrescenta Ferreira, respostas baseadas em informações incompletas tendem a ser mais imprecisas. "Com poucas informações que o paciente imputa ali, a chance de não vir algo certo é muito alta."

O risco é agravado pelo tom convincente das respostas. "Independentemente de estar certa ou não, a resposta é convincente. O paciente pode acreditar nos dois terços errados como se fossem verdade."

Os exemplos do estudo ajudam a entender o impacto prático. Em um dos cenários, dois participantes relataram sintomas clássicos de hemorragia subaracnóidea —dor de cabeça súbita e devastadora, descrita como "a pior da vida".

Um recebeu a orientação correta de buscar atendimento de emergência imediato. O outro, após interação semelhante, foi aconselhado a deitar em um quarto escuro e repousar. Pequenas variações na forma de perguntar levaram a recomendações opostas.

Em outro caso, envolvendo sinais compatíveis com embolia pulmonar, os usuários tiveram dificuldade em distinguir, entre múltiplas hipóteses apresentadas, qual era a mais grave.

Ferreira ressalta que o problema não se restringe aos pacientes. "O letramento dos profissionais de saúde é quase próximo ao dos pacientes. Eles sabem nada ou quase nada", diz. Segundo ele, muitos médicos utilizam versões gratuitas de chatbots sem treinamento formal e, às vezes, inserem dados sensíveis sem considerar implicações legais.

"Esse médico também não é uma pessoa apta a validar o output [informação adequadamente formatada] dali. Ele está um pouco melhor que o paciente, mas não muito melhor." Para ele, gestores de saúde deveriam investir em capacitação e em ferramentas institucionais, sob risco de ampliar erros e litígios.

Para Alexandre Chiavegatto, professor de inteligência artificial da Faculdade de Saúde Pública da USP, a pesquisa utilizou algoritmos já ultrapassados, mas levanta uma questão mais ampla.

"Eu e a maioria dos pesquisadores da área somos contra testar IA para uso direto por pacientes. Os algoritmos devem auxiliar médicos e outros profissionais, e passar por um filtro de profissionais treinados antes de chegar ao paciente."

Chiavegatto lembra que o uso já é uma realidade, mesmo com sistemas ainda em estágio inicial. "Estamos na infância do potencial da IA na saúde. A área precisa ser tratada como algo experimental e em construção."

Ao mesmo tempo, ele vê na tecnologia um potencial transformador. "O maior benefício da IA será na saúde, principalmente para auxiliar decisões em regiões remotas do Brasil, onde não há especialistas."

Segundo ele, médicos desses locais poderão ter acesso a ferramentas de apoio semelhantes às dos grandes centros, com potencial de reduzir desigualdades históricas no acesso ao cuidado.

O estudo também mostrou que os modelos funcionam melhor em provas do que na vida real. Em um banco de questões amplamente utilizado em exames médicos, o MedQA, os sistemas superaram com folga a marca de 60% considerada suficiente para aprovação.

No entanto, o alto desempenho nesse teste não se correlacionou com bons resultados nas interações com pessoas reais. Simulações com "pacientes virtuais" —em que uma IA conversa com outra— também falharam em prever os problemas observados com humanos.

Para os autores, o ponto crítico está na interface humano-máquina. Muitos participantes forneceram informações incompletas; outros ignoraram recomendações corretas. Alguns pareceram atribuir características humanas à ferramenta e basearam sua decisão na confiança transmitida pelo texto.

<https://www1.folha.uol.com.br/eqilibrioesaude/2026/02/ia-na-saude-vai-bem-na-teoria-mas-falha-com-pacientes-reais-mostra-estudo.shtml>

Veículo: Online -> Portal -> Portal Folha de S. Paulo