



Publicado em 06/03/2026 - 10:16

ChatGPT funciona para emergências médicas? Estudo lista falhas, viés racial e dúvidas sobre uso da IA

Estudo mostra que IA recomendou esperar até 48 horas por consulta em 51,6% das emergências reais e foi influenciada por comentários de familiares sobre os sintomas.

Por Lara Cáfaros*, g1

Um estudo que avaliou o uso do ChatGPT Health para analisar sintomas e exames mostrou que a ferramenta recomendou um nível de cuidado menor do que o necessário em mais da metade das emergências. A pesquisa também identificou viés racial nas respostas e influência de comentários de familiares nas recomendações.

A pesquisa foi realizada por médicos e cientistas da Escola de Medicina Icahn no Monte Sinai, em Nova York, foi publicado na revista Nature, uma das mais conceituadas do mundo. Em entrevista ao g1, Ashwin Ramaswamy, pesquisador responsável pelo estudo, afirma que a principal preocupação é que os erros de diagnósticos feitos por IA aparecem justamente nos casos mais graves.

ChatGPT Health: é uma ferramenta de saúde voltada para o consumidor, lançada recentemente pela OpenAI. Foi projetada para fornecer orientação de saúde diretamente ao público, tendo como função principal recomendar a urgência com que um usuário deve procurar um médico após relatar seus sintomas.

A pesquisa mostrou que a ferramenta recomendou um nível de cuidado menos urgente do que o necessário em 51,6% das emergências reais, sugerindo que pacientes em estado crítico esperassem de 24 a 48 horas por uma consulta médica, em vez de irem ao pronto-socorro.

“O sistema recomendou ‘monitorar em casa’ para um homem negro em cetoacidose diabética (complicação grave do diabetes), mas ‘ir ao pronto-socorro agora’ para um homem branco com a mesma apresentação clínica. Em um país como o Brasil, onde desigualdades raciais na saúde já são uma preocupação séria, esse tipo de inconsistência precisa ser analisado com atenção”, explica Ashwin.

Qual o perigo de levar em consideração o “diagnóstico” feito por IA?

Em mais da metade das vezes em que um paciente apresentava uma condição grave, a inteligência artificial falhou em recomendar atendimento imediato em um pronto-socorro. Em vez disso, sugeriu que o usuário procurasse um médico apenas entre 24 e 48 horas depois.

Com esse resultado, surge uma dúvida: consultar uma IA seria o mesmo que jogar uma moeda e contar com a sorte? Para o coordenador do estudo, a resposta é mais complexa.

“Uma moeda é aleatória, e as pessoas sabem que não devem confiar em uma moeda. O ChatGPT Health erra de forma seletiva. Ele acerta casos de gravidade média em mais de 90% das vezes, o que gera confiança. Depois falha justamente nos casos que mais importam — emergências reais — em mais da metade das vezes. Esse padrão é mais perigoso do que um cara ou coroa, porque o usuário não tem como saber quando o sistema passou de confiável para não confiável. Você ganha confiança depois de dez boas respostas, e a décima primeira diz para você ficar em casa durante uma crise diabética”, afirma.

Quando mais dados não ajudam

A inclusão de exames e outros dados médicos objetivos, em teoria, deveria melhorar a precisão do diagnóstico. No entanto, o estudo mostrou que, em alguns casos, essas informações podem criar uma falsa sensação de segurança.

Segundo os pesquisadores, a IA parece se “tranquilizar” ao encontrar alguns valores normais em meio a um quadro grave. O sistema tende a se fixar em resultados que parecem positivos isoladamente — como um nível normal de potássio ou uma creatinina dentro da faixa esperada — e acaba ignorando o quadro clínico geral que indicaria uma emergência.

Um médico, por exemplo, olha para um conjunto de fatores, como pH levemente alterado (indica um pequeno desequilíbrio na acidez ou alcalinidade em uma região do corpo), glicemia de 320 (considerada muito alta e indica um estado de hiperglicemia severa ou diabetes descompensado) e bicarbonato baixo (geralmente indica uma condição onde o corpo produz muito ácido ou perde muito bicarbonato), e reconhece uma cetoacidose diabética. Já o sistema de IA às vezes foca no que parece normal em vez do que é perigoso.

Por outro lado, a inclusão de achados objetivos aumentou a precisão geral das recomendações de 54,6% para 77,9% em casos leves e não urgentes.

Influência de amigos ou familiares

O estudo também mostrou que a presença de comentários de familiares ou amigos pode influenciar a recomendação da IA.

Quando os pesquisadores incluíram frases de pessoas próximas minimizando os sintomas, como “acho que você está bem” ou “provavelmente não é nada”, o ChatGPT Health se tornou quase 12 vezes mais propenso a recomendar um nível menor de atendimento.

Segundo Ashiwn, esse comportamento reflete o que os pesquisadores chamam de viés de ancoragem, um problema conhecido no raciocínio clínico humano. “A diferença é que médicos são treinados para reconhecer e resistir a esse viés”, explica.

Riscos em situações graves

O estudo também indica que o sistema de segurança do ChatGPT Health não funciona de forma lógica ou previsível em situações de risco de vida.

“Um recurso de intervenção em crise suicida que funciona 100% das vezes em um contexto e 0% em um contexto quase idêntico não é um mecanismo de proteção — é uma loteria. Reguladores deveriam exigir que recursos críticos de segurança funcionem de forma confiável em cenários clinicamente equivalentes, e que isso seja verificado de forma independente”, ressalta o autor do estudo.

Metodologia

A pesquisa utilizou um experimento fatorial estruturado, que estuda simultaneamente dois ou mais fatores e seus níveis, combinando todos entre si para analisar efeitos principais e interações.

A equipe de especialistas criou 60 casos clínicos fictícios baseados em diretrizes médicas reais de 58 sociedades profissionais. Esses casos cobriram 21 áreas da medicina e foram divididos em dois tipos: apresentações apenas com sintomas relatados pelo paciente e versões que incluíam dados objetivos, como resultados de exames laboratoriais e sinais vitais.

O diferencial da metodologia foi o chamado “design fatorial”. Os pesquisadores não apenas perguntaram sobre a doença, mas criaram 16 variações para cada caso, alterando propositalmente fatores não clínicos para observar se a IA mudaria de opinião. Entre as variáveis testadas estavam:

- Raça e Gênero: Pacientes negros vs. brancos; homens vs. mulheres.
- Ancoragem: Inclusão de comentários de amigos ou familiares que minimizavam a gravidade da situação.
- Barreiras de Acesso: Menção a dificuldades de transporte ou falta de seguro saúde

É preciso regular a inteligência artificial na saúde?

Segundo Antônio Carlos, coordenador da Comissão de Saúde Digital da Associação Médica Brasileira (AMB), o Brasil já possui algumas bases regulatórias aplicáveis, como a Lei Geral de Proteção de Dados (LGPD), que trata de dados sensíveis, além da atuação da Autoridade Nacional de Proteção de Dados (ANPD) e das regras da Anvisa quando o software tem finalidade médica.

“Mas precisamos avançar em critérios claros para nível de risco. Informação geral é uma coisa; triagem, diagnóstico e apoio à decisão exigem validação, monitoramento, transparência e governança”, diz.

“Não deveríamos estar em uma situação em que pesquisadores acadêmicos precisam correr para avaliar um produto depois que dezenas de milhões de pessoas já estão usando o ChatGPT para perguntas de saúde toda semana. Acreditamos que a avaliação independente de segurança deveria ser uma etapa rotineira antes que esses produtos cheguem ao público, da mesma forma que não pularíamos testes de segurança para um novo medicamento apenas porque ele parece promissor”, pontua Ashwin.

Quando a inteligência artificial pode ajudar na saúde?

De acordo com os pesquisadores, ferramentas de IA podem ser úteis para entender melhor um diagnóstico, pesquisar efeitos colaterais de medicamentos ou esclarecer dúvidas simples.

A recomendação, porém, é que elas sejam usadas como complemento ao médico, e não como substituto.

“Em resumo, a IA pode ajudar a organizar informações e orientar o paciente, mas não deve tomar decisões sozinha. Substituir consulta, diagnóstico ou prescrição são limites muito claros. A IA não faz exame físico, não mede sinais vitais e não responde legalmente pelo cuidado. Por isso, não pode assumir o papel do médico”, conta o coordenador da AMB.

Busque ajuda

Em caso de sintomas de depressão, procure a ajuda de um profissional habilitado. O Ministério da Saúde divulga os seguintes endereços para ajuda:

- CAPS e Unidades Básicas de Saúde (Saúde da Família, Postos e Centros de Saúde).
- UPA 24H, SAMU 192, Pronto Socorro; Hospitais
- Centro de Valorização da Vida – 188 (ligação gratuita). O CVV – Centro de Valorização da Vida realiza apoio emocional e prevenção do suicídio, atendendo voluntária e gratuitamente todas as pessoas que querem e precisam conversar, sob total sigilo, por telefone, e-mail, chat e voip, 24 horas por dia, todos os dias. A ligação para o CVV em parceria com o SUS, por meio do número 188, é gratuita a partir de qualquer linha telefônica fixa ou celular.
- Também é possível acessar www.cvv.org.br para chat, Skype, e-mail e mais informações sobre a ligação gratuita.

*(Estagiária, sob supervisão de Ardilhes Moreira).

<https://g1.globo.com/saude/noticia/2026/03/06/chatgpt-emergencias-medicas-estudo-lista-falhas-vies-racial-e-duvidas.ghtml>

Veículo: Online -> Portal -> Portal G1