IA contra *fake news*

Pesquisadores de universidades brasileiras desenvolvem aplicações para combater desinformação

SARAH SCHMIDT __ ilustrações **JOÃO MONTANARO**

Em março, um vídeo da suposta explosão de uma bomba no porto de Haifa, em Israel, chegou à caixa de e-mail do Laboratório de Inteligência Artificial da Universidade Estadual de Campinas (Unicamp), o Recod.ai. Os remetentes eram jornalistas da agência internacional de notícias *France-Presse* (*AFP*), que pediam ajuda para verificar a autenticidade do material. A requisição foi intermediada pela Witness, organização internacional que ajuda pessoas a produzir e divulgar vídeos para defender direitos humanos e mantém parceria com instituições de pesquisa para verificar se os conteúdos são reais ou manipulados.

Coube ao cientista da computação Mateus de Padua Vicente, estudante de doutorado da Unicamp, analisar o vídeo do ataque usando diferentes ferramentas. A conclusão foi que o material provavelmente era falso. Seu relatório identificou *frames* – momentos estáticos de vídeo – com alta probabilidade de serem artificiais. "Nosso modelo apontou, por exemplo, que a trilha de fumaça parecia suave demais, sem textura, e que os prédios ao fundo estavam borrados. Outro ponto: a iluminação na fumaça não parecia bater com o ponto focal da luz do sol. O momento da explosão indicou que ela não parecia real, pois ficou borrada justamente quando acontecia", explica Vicente. Por ora, a ferramenta só permite a análise de imagens e, por isso, é preciso analisar quadro a quadro. Ele observa que esses casos concretos de checagem funcionam como um "laboratório vivo" para testar e melhorar as ferramentas. "Os casos reais expõem limitações que não aparecem em bases de dados disponíveis ou referenciais acadêmicos e são muito úteis para avaliar a capacidade das ferramentas de gerar dados precisos."

O Recod.ai é um dos laboratórios de universidades brasileiras que têm produzido ferramentas para identificar conteúdos falsos e combater a desinformação em diferentes esferas, da política à saúde. De acordo com o cientista da computação Anderson Rocha, coordenador do Recod.ai e orientador de Vicente, nos primeiros meses deste ano já foram analisados cerca de 10 vídeos em parceria com a Witness. "Aparentemente, está havendo um aumento na disseminação desse tipo de conteúdo. No ano passado inteiro, recebemos 20 vídeos suspeitos de manipulação; alguns envolviam políticos ou autoridades em cargos importantes", observa. Essa percepção coincide com uma tendência apontada em um relatório da agência Lupa, especializada em checagem de informações, divulgado em fevereiro, mostrando que conteúdos falsos gerados por inteligência artificial (IA) cresceram 308% entre 2024 e 2025 e passaram a representar 25% das verificações da empresa no país.

O grupo da Unicamp coordena um projeto temático apoiado pela FAPESP, o Horus, que busca criar técnicas de IA para detectar e analisar imagens sintéticas. Um dos modelos que integram o conjunto de ferramentas desenvolvidas pela equipe é o FakeScope, criado em parceria com pesquisadores chineses para identificar imagens geradas por IA e, ao mesmo tempo,

explicar por que elas são falsas. Sua funcionalidade foi apresentada em um texto em *preprint* (sem revisão por pares) publicado em março de 2025 no repositório arXiv. "Diferentemente de métodos que apenas classificam imagens como reais ou falsas, o FakeScope foi treinado com grandes conjuntos de imagens geradas por IA e reais, além de anotações detalhadas sobre indícios visuais, como luz, textura e bordas", explica Vicente.

Outro modelo é o Pixel-Inconsistency, que analisa inconsistências nos pixels de imagens e vídeos para identificar possíveis manipulações. No caso do vídeo sobre a suposta explosão em Israel, foi essa ferramenta que destacou as irregularidades na fumaça, possivelmente adicionadas por IA. Há, ainda, o Deepfake Detection System, ferramenta que analisa biometria facial para identificar manipulações em vídeos que mostram pessoas conhecidas. Um dos casos enviados pela Witness envolvia um vídeo em que o presidente do Líbano, general Joseph Aoun, supostamente sugeria que planejava criminalizar a organização política paramilitar islâmica Hezbollah, instalada há quatro décadas no país, responsável por embates com Israel. O vídeo era falso. Sinais como mudanças no modelo do relógio, movimentos faciais pouco naturais e distorções no fundo foram apontados pelo DeepFake Detection, enquanto o FakeScope identificou distorções em uma bandeira na sala e problemas nas bordas da imagem. "Além da parceria com a Witness, também já ajudamos em casos de checagem de fatos em parceria com veículos brasileiros, como o G1 e a Lupa", conta Rocha. Para o período de eleições deste ano, o laboratório deve formalizar uma parceria com a agência. Já a ferramenta de detecção de deepfake foi cedida para ser usada pelas polícias legislativas da Câmara e do Senado.

A desinformação em saúde, especialmente a relacionada com vacinas, também é alvo de pesquisas do laboratório. No projeto Aletheia, financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e pelo Ministério da Saúde, os pesquisadores mapearam 119 canais e grupos públicos em português e inglês no aplicativo de mensagens Telegram que promoviam narrativas antivacina e anticiência entre 2020 e 2025, reunindo uma base de dados com cerca de 4 milhões de postagens, entre imagens, textos, vídeos e documentos. "Vimos que vários grupos que estavam no WhatsApp migraram para o Telegram por ser um espaço menos monitorado", observa Rocha.

O banco de dados aberto está disponível em um repositório da Unicamp para pesquisadores que queiram produzir estudos sobre como esse tipo de desinformação se espalha e pode influenciar o comportamento público. "Estamos analisando as mensagens para entender o comportamento dos usuários e como os conteúdos circulam nesses espaços", explica a jornalista Ana Carolina Monari, pesquisadora em estágio de pós-doutorado no projeto. "Desenvolvemos uma taxonomia de discursos de desinformação em saúde, com 17 categorias, que servirá para treinar modelos de IA em português que ajudem a identificar automaticamente os problemas nas redes sociais", complementa. Para o levantamento, os pesquisadores desenvolveram um modelo que pudesse identificar as postagens desinformativas sobre vacinação.

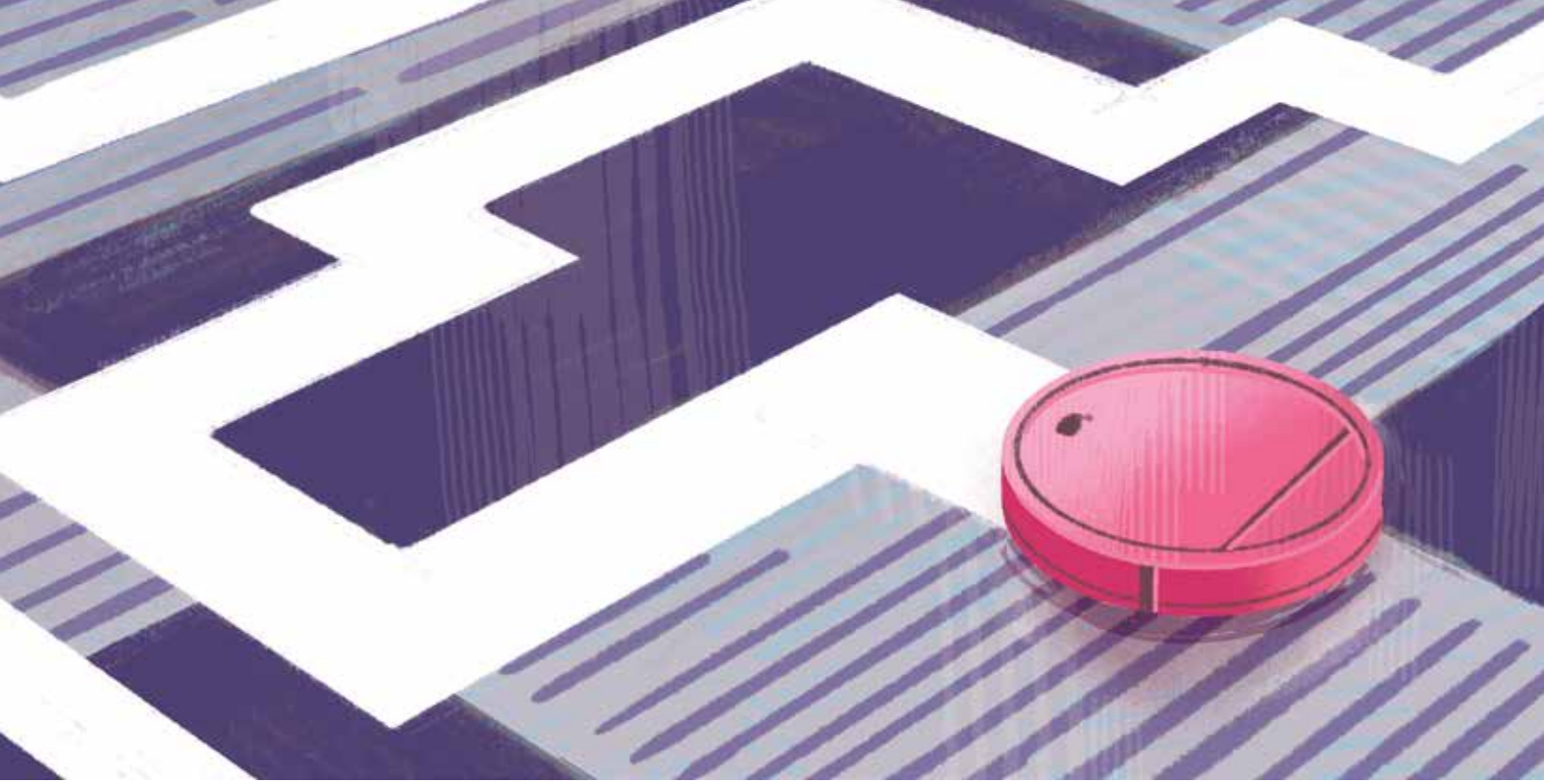
WHATSAPP E OUTROS MODELOS

Um programa para o WhatsApp que usa inteligência artificial para checar informações foi desenvolvido por três estudantes de ciência da computação da Universidade de São Paulo (USP) em São Carlos. O usuário precisa apenas salvar em sua lista de contatos o número de telefone vinculado ao programa e encaminhar para ele mensagens suspeitas. "Nosso slogan é 'verificar é tão fácil quanto encaminhar'. Pode ser texto, vídeo, áudio ou link, e o software analisa o conteúdo", conta o estudante Cauê Paiva Lira, de 21 anos, um dos criadores da ferramenta Tá Certo Isso AI?.

Ele conta que a ideia surgiu em uma maratona que reuniu programadores na USP para criar soluções de IA contra a desinformação, na qual a proposta acabou sendo vencedora. "Fomos selecionados por um programa internacional chamado AI for Good, organizado pela Brazil Conference e realizado no MIT [Instituto de Tecnologia de Massachusetts] e na Universidade Harvard, entre mais de 170 inscritos", conta Lira, que se preparava com o grupo para apresentar o projeto em Boston, nos Estados Unidos, no final de março.

Casos concretos de checagem ajudam a testar e a melhorar as ferramentas

Os estudantes usaram modelos de linguagem como o Gemini e o GPT, mas, ressalta Lira, com



um diferencial importante: o chatbot, assistente virtual que simula conversas humanas, faz a verificação com base exclusivamente em fontes confiáveis, como agências de checagem. Um exemplo foi o caso de uma mensagem que dizia que era preciso pagar pela CNH Social no estado do Paraná. A conclusão do aplicativo é de que a informação era falsa, já que a emissão da carteira de motorista é gratuita para pessoas de baixa renda. A ferramenta checou a informação em reportagens do G1 e do jornal *O Estado de S. Paulo*.

Quando não consegue verificar a autenticidade da informação, o aplicativo avisa para o usuário. Depois de serem analisadas, as mensagens são anonimizadas, para evitar a exposição dos usuários, e alimentam uma base de dados pública, que pode ser usada por outros pesquisadores. "Agora buscamos expandir o projeto, formar parcerias e nos aproximar mais da pesquisa acadêmica, porque acreditamos que combater a desinformação com IA é uma área com muito potencial de impacto social", diz o estudante.

Na Universidade Federal de São Carlos (UFSCar), um grupo de pesquisadores do Laboratório Interfaces desenvolveu um modelo que busca detectar notícias falsas de uma forma diferente da maioria dos modelos tradicionais. "O algoritmo aprende a identificar *fake news* tendo acesso apenas a exemplos de conteúdos falsos, para encontrar padrões mesmo sem conhecer todos os casos possíveis", explica o matemático Guilherme Henrique Messias, primeiro autor de um artigo que explica o modelo, publicado em 2026 pela editora Springer nos anais da *Brazilian Conference on Intelligent Systems*.

"A nossa proposta foi desenvolver um modelo que funcione mesmo com poucos dados", diz o cientista da computação Alan Valejo, da UFSCar, orientador de Messias no doutorado em ciência da computação. "Em vez de precisar de grandes bases com notícias verdadeiras e falsas, o modelo aprende usando apenas exemplos de notícias falsas, o que resolve um dos principais desafios da área: a dificuldade e o alto custo de rotular dados confiáveis", explica.

Ele funciona transformando o conteúdo de textos em representações gráficas para detectar características típicas de *fake news*. Embora ainda não esteja disponível como um software funcional, sua base está no GitHub, plataforma colaborativa de projetos de software, e pode ser usada por jornalistas ou pesquisadores com conhecimentos básicos de programação em Python. "Além de identificar desinformação, o modelo pode ser usado em outras áreas com pouca informação disponível, como na análise de redes sociais e científicas, para encontrar padrões e relações escondidas. De forma geral, o programa ajuda a descobrir informações importantes mesmo quando há poucos dados organizados ou completos", vislumbra Messias. ●

Os projetos e os artigos científicos consultados para esta reportagem estão listados na versão on-line.