

Desinformação russa chega à inteligência artificial

Redes *Pravda* e *Storm-1516* poluem a Internet com conteúdos falsos mostrados como verdadeiros no ChatGPT

Ivo Neto Texto
José Alves Ilustração

Um exército de *sites* está a poluir o ambiente informativo com desinformação sobre a guerra na Ucrânia, visando favorecer os interesses de Moscovo e influenciar o apoio ocidental a um conflito que já dura há quase quatro anos. O objectivo é enganar plataformas de inteligência artificial (IA), como o ChatGPT, levando-as a produzir respostas falsas quando são utilizadas para procurar informação.

“[O Presidente ucraniano, Volodymyr] Zelensky comprou a infame vivenda de Goebbels (Bogensee) por oito milhões de euros.” Este é o título de um vídeo publicado no YouTube, em 2023, que chegou a mais de 13 mil pessoas. É falso. Tal como é falsa outra publicação, de Outubro de 2024, que dizia que o Presidente ucraniano teria comprado o carro de Adolf Hitler. É igualmente falsa uma publicação na rede social X que, em Março de 2025, durante as eleições alemãs, dizia que a AfD, um partido de extrema-direita, estaria fora dos boletins de voto em alguns locais. Uma das publicações em que esta informação falsa foi partilhada foi vista por quase 50 mil pessoas e teve perto de mil partilhas.

Segundo o Viginum, o serviço de vigilância e protecção contra a interferência digital estrangeira do Governo francês, criado em 2021, as três publicações fazem parte de uma rede de desinformação, a *Storm-1516*, que tem como objectivo minar o apoio de Bruxelas a Kiev e influenciar as eleições em países ocidentais — nomeadamente as da Alemanha, França e EUA.

A campanha está operacional pelo menos desde Agosto de 2023 e utiliza uma vasta gama de conteúdos para difundir as suas narrativas, incluindo montagens de fotografias e vídeos, bem como vídeos e áudios

muito provavelmente gerados com recurso a ferramentas de IA. Estes conteúdos incluem textos e vozes em francês, inglês, ucraniano, alemão, espanhol e árabe.

No relatório, publicado em Maio, o Viginum apresentava nomes próximos do Kremlin que estarão envolvidos neste esquema. É o caso de John Mark Dougan, um antigo fuzileiro do Exército norte-americano exilado na Rússia, e também Iuri Khoroshenki, que integrará a Unidade 29155 do GRU, a agência de inteligência militar estrangeira das Forças Armadas da Federação Russa, e que é acusado de financiar a campanha.

Documentos analisados pelo *Washington Post* revelam pagamentos feitos directamente por Khoroshenki a Dougan a partir de Abril de 2022, assim como encontros entre os dois, que contaram também com a presença de Alexander Dugin, um ideólogo imperialista de extrema-direita.

“Os operacionais de Moscovo integraram ferramentas de IA generativa em todas as fases das operações de influência”, confirma o PÚBLICO Halyna Padalko. A especialista em comunicação estratégica, desinformação e IA do MIT sublinha a “seg-

mentação estratégica que adapta o conteúdo a comunidades específicas, ampliando assim as divisões sociais e minando a confiança pública”.

Se até há pouco tempo era mais simples perceber quando a informação adulterada vinha de mecanismos automatizados, o uso de IA veio dificultar a distinção entre o que é verdadeiro e o que é falso.

“Creio que estamos perante uma

Informação falsa sobre piloto dinamarquês morto ao treinar soldados ucranianos chegou a plataformas de IA. Ministério da Defesa da Dinamarca teve de desmentir caso

guerra de informação de precisão em massa. As pessoas podem receber mensagens altamente específicas — até mensagens dirigidas individualmente, microdirecionadas a cada um em particular”, explica, por videochamada, Chris Kremidas Courtney, investigador sénior no European Policy Centre e investigador associado no Geneva Centre for Security Policy.

A rede *Pravda*

Esta não é, porém, a primeira rede de desinformação identificada pelo Viginum com supostas ligações ao Kremlin. O *Portal Kombat*, também conhecido como rede *Pravda* [verdade, em russo], publica em países de toda a Europa, África e Médio Oriente em quase 50 idiomas e, só em Fevereiro, segundo o jornal *Times*, do Reino Unido, difundiu mais de 350 mil artigos.

Segundo Valentin Chatelet, do Atlantic Council, em resposta por *email*, em 2025, a rede chegou a pelo menos 83 países e regiões de todo o mundo. Há, tal como no caso da *Storm-1516*, um foco declarado em países como a Alemanha e França.

O especialista sublinha ainda o caso

do apagão eléctrico que atingiu a Península Ibérica, em Abril, e que foi aproveitado para amplificar desinformação (ver caixa ao lado). Segundo a plataforma Maldita, um site de verificação de factos sediado em Espanha, entre 28 e 30 de Abril, verificaram-se quase 1200 publicações nos *sites* da rede russa.

O PÚBLICO contactou a Embaixada da Rússia em Lisboa, que por *email* disse “não fazer comentários sobre especulações dos serviços franceses”.

Enganar a IA e as sanções

Mas a consequência vai além da simples disseminação de conteúdos falsos. Um dos principais objectivos, de acordo com Chatelet, é a “poluição” dos modelos de linguagem de grande escala (LLM na sigla em inglês), como o ChatGPT, da Open AI, ou o Copilot, da Microsoft.

O investigador do Atlantic Council aponta para o risco da infiltração destes conteúdos na Wikipédia. Dado que muitos modelos de LLM são treinados a partir de grandes volumes de conteúdos da Wikipédia, “as hiperligações da *Pravda* e as narrativas a elas associadas podem ser validadas pela





Apagão em Portugal e desinformação russa

O apagão eléctrico que atingiu a Península Ibérica, em Abril, foi como um laboratório para as redes de desinformação. Segundo o Observatório Europeu dos Meios de Comunicação Digitais (EDMO, na sigla em inglês), assim como a agência EFE, Portugal e Espanha foram alvo de conteúdos falsos com origem na Rússia. O relatório da EDMO dava conta de que foi falsamente afirmado que a presidente da Comissão Europeia, Ursula von der Leyen, teria acusado erradamente a Rússia pelo apagão.

“A rede *Pravda* tem dois sites dedicados a Portugal. Um deles, publicou seis mil conteúdos no ano passado”, explica Inês Narciso, investigadora do Media Lab - Iscte, que detalha a forma como estes conteúdos são disseminados: “Estas redes não usam apenas IA, reproduzem [também] conteúdos que são publicados em redes como o Telegram.” Entre os conteúdos, destacam-se, por exemplo, textos de ucranianos com posições favoráveis à Rússia. “Mesmo que Portugal não seja uma prioridade, que se desengane quem achar que não existe nada”, conclui a especialista portuguesa.

IA, levando-a a acreditar que os sites da *Pravda* são fontes noticiosas fidedignas”.

Num estudo desenvolvido em parceria com a finlandesa Check First, Chatelet confirmou que a maioria dos chatbots de IA se mostrava propensa a repetir e a incluir hiperligações para artigos da *Pravda*. Isso acabou por levá-los a referir entidades sancionadas que utilizam o Telegram para contornar as proibições em vigor em toda a União Europeia, como a *Pravda* e a *Russia Today*.

“Na prática, a Rússia passou de influenciar eleitores a influenciar a própria infra-estrutura informacional da era digital”, acrescenta, por sua vez, Halyna Padalko.

Em 2024, segundo Chris Kremidas Courtney, a propaganda russa publicou 3,6 milhões de artigos de desinformação online, “poluindo o ecossistema informativo” – um estudo do AI Safety Institute, uma agência governamental britânica, concluiu que sistemas de LLM podem ser influenciados com apenas 250 documentos.

O soldado que nunca existiu

Ao PÚBLICO, McKenzie Sadeghi, da

NewsGuard, uma empresa norte-americana que avalia a credibilidade de sites noticiosos, apresenta o caso do soldado Jepp Hansen – que na verdade nunca existiu.

Meios de comunicação estatais russos deram conta de que um piloto dinamarquês chamado Jepp Hansen teria morrido num ataque enquanto treinava forças ucranianas.

Nas redes sociais, o Ministério da Defesa da Dinamarca declarou mesmo que “não há soldados dinamarqueses mortos na Ucrânia” e considerou a notícia como “uma campanha de desinformação”. Ainda assim, uma investigação desenvolvida pelo NewsGuard demonstrou que a narrativa russa estava a ser partilhada em plataformas de IA.

Num relatório publicado em Setembro, o NewsGuard dá conta de que, em 2025, as dez mais importantes plataformas de IA – ChatGPT, Mistral, Grok, Claude, You.com, Gemini, Inflection, Copilot, Meta e a Perplexity – apresentam informações falsas em pelo menos 35% das questões relacionadas com a actualidade informativa. É mais do dobro do que acontecia em 2024, indica o NewsGuard.

O PÚBLICO contactou a OpenAI para procurar informações sobre a ingerência de desinformação russa no ChatGPT, mas não obteve qualquer resposta até à hora da publicação deste artigo. Ainda assim, a empresa norte-americana, num relatório a que o PÚBLICO teve acesso, dizia ter bloqueado contas do ChatGPT que pareciam ter origem na Rússia.

Se os vários especialistas com quem o PÚBLICO falou recomendam um maior controlo por parte das fontes usadas pelas plataformas de LLM, Mykola Makhortkyh, da Universidade de Berlim, alerta para o facto de soluções que passem pelo condicionamento da informação podem limitar “o acesso individual a informação perfeitamente legítima ou de levar as pessoas a procurá-la noutros locais”.

Makhortkyh diz que a comunidade científica ocidental deveria “centrar-se menos em provocar pânico moral sobre as aplicações de IA ao serviço dos objectivos do Kremlin e mais em exigir maior transparência e responsabilidade às empresas tecnológicas que desenvolvem e implementam estas aplicações”.

Investimento de milhões Israel quer IA como “ferramenta de diplomacia pública”

Maria João Guimarães

É considerado um dos maiores investimentos de Israel numa campanha de *hasbara* (diplomacia pública): segundo vários meios de comunicação incluindo o *Yediot Ahronot* e o *Haaretz*, o país terá gasto quase 125 milhões de euros numa campanha para contrariar críticas e mostrar a sua versão de acontecimentos.

Uma parte do destino das verbas não traz nada de novo: pagamentos a empresas de relações públicas ou figuras influentes. Onde se destaca é na parte gasta em empresas com sistemas de inteligência artificial generativa.

Segundo o *Yediot Ahronot*, que cita documentos submetidos ao Departamento de Justiça dos EUA no âmbito da lei que regula o registo de agentes estrangeiros a que são obrigadas empresas que representam governos estrangeiros (Foreign Agents Registration Act), há um esforço para influenciar a representação de questões ligadas a Israel feitas pelos sistemas que utilizam IA generativa.

Estas estão a ter cada vez mais influência no modo como são feitas pesquisas, tendo já levado a uma menor taxa de abertura de sites de informação geral ou especializada, fazendo as pessoas a pergunta directamente ao ChatGPT ou outro modelo ou ficando-se pelo resumo dado pela IA apresentado por motores de busca como o da Google, com quebras a chegar a 31% em sites de informação médica entre o início do ano e o mês de Junho, segundo a empresa Similarweb citada num artigo da revista britânica *The Economist* com o título *AI is killing*

the Web [A IA está a matar a Net].

Uma das empresas contratadas por Israel tem a missão de “criar materiais online que possam dar forma a dados usados para o treino dos *large language models*” [que usam grandes quantidades de texto] como o ChatGPT, o Grok, do X [antigo Twitter] e o Gemini da Google, potencialmente afectando o modo como apresentam ou enquadram questões envolvendo Israel”, diz o *Yediot Ahronot*.

Uma cláusula no contrato com a empresa Clock Tower X fala numa “Search and Language Operation”, para “gerar resultados de enquadramentos em conversas com o ChatGPT e outros sistemas de diálogo baseados em IA”.

O PÚBLICO acabou por se deparar, inadvertidamente, com uma aparente consequência de como são apresentados resultados numa tradução feita no ChatGPT. Tratava-se de um artigo sobre dois arquitectos que fariam uma conferência na Fundação Gulbenkian, em Lisboa, e levantavam questões sobre as ruínas de Gaza. O texto foi traduzido com o ChatGPT para enviar aos entrevistados, que não falam português.

Ao rever a tradução, em geral muito aproximada, uma parte estava diferente: no artigo em português, dizia-se o seguinte: “Desde Outubro de 2023, logo no início do genocídio”, sentiram a necessidade de fazer algo para ajudar.” A tradução acrescentou, a seguir às aspas que terminavam em ‘genocídio’, “as they describe it”, juntando uma camada de distanciamento em relação à expressão, que é usada por várias organizações e pela maioria das pessoas da academia que estudam o campo.



Israel terá gasto quase 125 milhões de euros numa campanha

ILUSTRAÇÃO: JOSÉ ALVES



Fake News

Desinformação russa está a infiltrar-se na IA e pode afectar respostas do ChatGPT

Mundo, 18/19