What's Next? Quatro Tendências a Considerar Para o Futuro da Desinformação

Inês Narciso[1]

Resumo: Este capítulo explora quatro tendências das desordens informativas. A primeira é a centralidade do indivíduo, quer pela relevância do conteúdo pessoal e crescimento da figura do influenciador, quer pela maturação de exércitos digitais, sobretudo na era pós-pandemia. A segunda aprofunda a natureza subtil das campanhas de desinformação modernas, ancoradas nas alterações de contexto e nas pequenas nuances que querem dar ao utilizador a ideia de construção de um puzzle que é ele que vai desvendar. A terceira revela como a desinformação tem migrado para plataformas de mensagens clandestinas, longe do escrutínio das redes sociais tradicionais. Por último, a quarta explora a forma como a desinformação se tornou autossuficiente e pronta a utilizar, graças ao desenvolvimento da inteligência artificial.

Quando procuramos ver além do horizonte no contexto das desordens informativas, temos de reconhecer que o seu futuro é esquivo e imprevisível. A era digital trouxe desafios sem precedentes, tornando cada vez mais difícil antecipar a extensão total da evolução da desinformação. No entanto, ao examinarmos algumas tendências atuais, quatro das quais são discutidas neste capítulo, podemos obter informações valiosas que servem como ponto de partida para projetar cenários futuros.

Ao contemplarmos o futuro da desinformação, é essencial reconhecer que surgirão novas tendências e tácticas, tornando as nossas previsões incompletas. No entanto, as quatro tendências destacadas neste artigo oferecem informações valiosas sobre o panorama atual e sublinham a necessidade de investigação contínua, colaboração e estratégias de adaptação.

[1] ISCTE-IUL.

Desinformação P2P

Tal como no marketing[2], a comunicação *peer to peer* tem vindo a ganhar espaço no ecossistema desinformativo. A confiança num produto, quer seja um novo artigo de beleza, quer seja uma narrativa desinformativa sobre a existência de chips nas vacinas, aumenta quando nos é sugerido por alguém próximo. A figura do *influencer*, com quem sentimos proximidade, mesmo que artificial, veio capitalizar esta relação de confiança. Em termos de investimento e recursos, uma única figura consegue distribuir uma narrativa a dezenas ou centenas de milhar de pessoas, o que é muito superior à reduzida capacidade de um perfil fabricado, com poucos seguidores e pouca tração.

Muitas vezes, estas figuras centrais são crentes absolutos nestas narrativas ou, pelo menos, alinham toda a sua base identitária nas mesmas, espelhando as profundas alterações que o ecossistema da desinformação sofreu, desde o famoso caso do Cambridge Analytica. Embora não tenham desaparecido as campanhas com recurso a centenas de perfis falsos que automática e deliberadamente partilham desinformação, o espaço é agora sobretudo tomado por exércitos digitais orgânicos, compostos por poucas pessoas, extremamente motivadas. A atual definição de desinformação da UE [12], que a apresenta sobretudo como algo que é intencionalmente criado e partilhado por organizações hostis com o intuito de enganar, peca por deixar para segundo plano o poder da mobilização individual autêntica a que assistimos na atualidade.

Já em 2021, um relatório do Center for Countering Digital Hate (CCDH) [3], destacava que 12 indivíduos eram responsáveis pela maior parte da desinformação e das teorias da conspiração anti-vacinas da Covid-19 que circulavam nas plataformas das redes sociais. Esta dúzia de *influencers* tinham um total de 59 milhões de seguidores em várias plataformas, sendo o Facebook a mais influente. O CCDH analisou mais de 800.000 mensagens e tweets no Facebook e Twitter e descobriu que 65% deles provinham destes 12 indivíduos. Só no Facebook, essa dúzia de *influencers* era responsável por 73% de todo o conteúdo anti-vacinas.

[2] O tamanho do mercado global de marketing de influenciadores mais do que duplicou desde 2019. Em 2023, o mercado foi estimado em um recorde de 21,1 mil milhões de dólares.

Em Portugal, uma *influencer* no TikTok com quase 2 milhões de *gostos* transitou para os ecrãs de televisão após convite para participar num *reality show*. Este convite foi feito apesar da mesma divulgar abertamente desinformação de saúde nas suas redes sociais, e de algumas alegações polémicas sobre vacinas e anti-depressivos terem contribuído para o crescimento da sua popularidade [6].

Para além disso, a inexistência de regulamentação deste mercado, ao contrário do mercado de anúncios pagos diretamente às plataformas, o qual já é alvo de relativo escrutínio desde as eleições americanas de 2016, torna-o também mais apetecível para campanhas de influência por atores hostis. Quando no ecrã do nosso telefone surge uma celebridade portuguesa a visitar um país do Médio Oriente, elogiando o país, não só como destino turístico, mas como defensor dos direitos das mulheres, importa compreender que estamos num mundo onde as redes sociais são, cada vez mais, uma arma de manipulação e interferência política.

Desinformação como peças de puzzle

No panorama de constante evolução da desinformação, uma das mudanças mais significativas é a sua transição de falsidades evidentes para manipulações subtis de contexto. A narrativa desinformativa não nos é apresentada como documentário factual, mas sim como filme de suspense, em que pequenas pistas, peças do puzzle, vão surgindo na nossa tela. Não nos é dito que as vacinas matam e têm chips no seu interior, mas são-nos apresentadas histórias de indivíduos vacinados que morreram. Não havendo prova de causalidade, tais ocorrências são apresentadas como coincidências que nos devem fazer questionar a segurança da vacinação.

Quem nunca se sentiu convencido que tinha sido o primeiro na sala de cinema a perceber quem era verdadeiramente o assassino? A crença numa teoria que é construída pelo utilizador, em que o mesmo junta diversas peças do puzzle e sozinho projeta a narrativa final é, em regra, superior a um cenário apresentado por terceiros. Vieses cognitivos como o efeito Dunning-Kruger, o viés de confirmação e os erros fundamentais de atribuição, todos contribuem para a emergência de teorias da conspiração como a QAnon [11]. Os defensores da teoria QAnon acreditam que existe uma rede de pessoas poderosas por todo

o mundo que, na verdade, seriam adoradores do Diabo, pedófilos e canibais e que utilizariam os seus recursos e influência para operar uma rede global de tráfico sexual de crianças. Uma das figuras acusadas de pertencer a esta rede é Hillary Clinton, candidata à presidência dos EUA em 2016.

A maior subtileza e dissimulação das narrativas desinformativas decorre também dos nossos sucessos, quer na literacia digital e capacidade de os utilizadores identificarem conteúdos evidentemente falsos, quer no trabalho das plataformas em implementar modelos de linguagem que identificam, de forma relativamente eficaz, conteúdo desinformativo. De facto, a infodemia que acompanhou a pandemia da Covid-19 em muito contribuiu para o desenvolvimento de melhores mecanismos de combate à desinformação, quer do ponto de vista dos utilizadores, quer no que diz respeito às plataformas. Em 2021, o Facebook removeu 26.4 milhões de publicações relacionadas com a Covid-19 que violavam os seus padrões da comunidade. Tais padrões foram, na maioria das plataformas, revistos e reforçados na ótica do combate à desinformação. Estas melhorias surgiram também em resultado da pressão da opinião pública e das instituições governamentais e de nova legislação, europeia e nacional, destacando-se o 'Digital Services Act'.

O estabelecimento de modelos de cooperação com as organizações de *fact-checking* e a introdução de rótulos identificando conteúdo enganador ou falso passou a ser mais comum. A identificação de conteúdo falso através de palavras-chave e a sua consequente sinalização ou remoção também aumentou. Esta capacidade foi amplamente potenciada pelo desenvolvimento da Inteligência Artificial (IA) e pelo uso de modelos de linguagem. Paralelamente, assistiu-se a um investimento maior em investigação e em projetos de literacia digital, como o 'Google News Initiative' e a 'Stop the Spread campaign'.

No passado era comum encontrar campanhas de desinformação baseadas em falsidades flagrantes e histórias fabricadas. À medida que os algoritmos baseados em IA e as ferramentas de verificação de factos melhoraram, o conteúdo evidentemente falso tornou-se mais suscetível de ser detetado. Nos últimos anos, assistiu-se à emergência de vários exemplos de desinformação mais subtil, aproveitando o poder do contexto e o uso de multimédia para enganar sem mentir abertamente.

Num estudo feito em Portugal sobre desinformação a circular no Facebook durante a campanha para as legislativas de 2019, uma das técnicas mais comuns identificadas pelos investigadores foi a descontextualização de conteúdo [2]. Uma notícia sobre o aumento da eletricidade de 2012, partilhada 7 anos depois e sendo atribuída ao governo então em funções, teve mais de 3 mil partilhas, e era visível nas centenas de comentários que a grande maioria dos utilizadores não tinha aberto a hiperligação, ou não se tinha apercebido que a notícia era antiga.

Um estudo recente evidencia como a indústria petrolífera tem vindo a liderar campanhas de desinformação contra o aquecimento global [9]. Nessa investigação, destaca-se como a mesma transitou da negação e contestação do aquecimento global, para campanhas que procuram, de forma mais subtil, mudar gradualmente perceções sobre o tema e foco, de forma a conseguir, discretamente, prosseguir a sua agenda.

Às mesmas conclusões chegaram investigadores britânicos quando começaram a analisar indícios de desinformação na Wikipédia [1]. A plataforma de conhecimento coletivo surge muitas vezes no topo das nossas pesquisas Google sobre um tema, e o nosso índice de confiança na informação contida na mesma é elevado. O desenvolvimento da IA, o forte investimento em moderadores humanos e o aperfeiçoamento dos algoritmos permite à plataforma ter uma elevada taxa de sucesso na identificação de desinformação. Mas a sua importância no mapeamento da opinião pública levou a que vários atores hostis aprendessem, segundo esta investigação, a fazer apenas pequenas alterações, incluindo o uso de certas palavras ou o reencaminhamento de links para, de forma dissimulada, conseguir desviar dos factos os leitores desta enciclopédia online.

Neste novo registo mais subtil da desinformação, somos levados a crer que chegámos a uma conclusão sozinhos, sem nos apercebermos das migalhas de pão enganadoras que nos conduziram a ela. O conteúdo também circula mais livremente nas plataformas, escapando à deteção. Embora o seu impacto direto possa ser menor, este conteúdo pode-se infiltrar mais facilmente no discurso público, sem ativar os mesmos sinais de alarme que as narrativas evidentemente falsas ativam.

Desinformação underground

À medida que as plataformas desenvolvem novas ferramentas para mitigação da desinformação, surgem também novas fronteiras à sua implementação: o mundo das plataformas de mensagens. Em 2022, as aplicações de mensagens eram responsáveis por 60% do tempo que os utilizadores estavam em aplicações 'sociais' [5]. O seu número de utilizadores também tem vindo a crescer a um ritmo superior ao das redes sociais. A principal consequência desta tendência para o combate à desinformação é que neste momento somos um lutador de olhos vendados. Com a proliferação destes espaços digitais fechados e a diminuição do acesso às API das redes sociais, o mapeamento da desinformação tornou-se cada vez mais difícil. À medida que as campanhas de desinformação encontram refúgio nestes cantos escondidos da Internet, aqueles que trabalham para as identificar e mitigar passam a ter como principal obstáculo o acesso aos dados.

As plataformas de mensagens como o WhatsApp e o Telegram tornaram-se, nos últimos anos, terreno fértil para a partilha de todo o tipo de conteúdo que as redes sociais foram banindo ou identificando como contrário aos seus padrões de comunidade. Isto inclui desinformação, a partilha de rumores ou boatos, discurso de ódio, racista, homofóbico ou outros e todo o tipo de conteúdo NSFW[3], desde vídeos grotescos de acidentes rodoviários, a pornografia, nomeadamente conteúdo multimédia íntimo partilhado sem consentimento (*revenge porn*). Encriptados e isolados, estes canais privados oferecem o ambiente perfeito para espalhar este tipo de conteúdos sem escrutínio, longe do olhar das autoridades e de quase impossível rastreamento forense à fonte de origem. Para um ator hostil que pretenda lançar uma campanha de influência, por exemplo numas eleições, utilizando desinformação, este é um meio de distribuição que lhe oferece muitas vantagens face às tradicionais redes sociais.

Quando identificada a campanha, é extremamente difícil avaliar o seu verdadeiro impacto. Nas redes sociais públicas, as ferramentas de rastreio e análise de dados fornecem informações valiosas sobre o alcance e o

[3] *Not safe for work*. É uma gíria utilizada na internet como indicação de alerta para conteúdos impróprios para serem visualizados em locais públicos ou no local de trabalho, como conteúdos violentos ou sexualmente explícitos.

envolvimento de conteúdos desinformativos. No entanto, estas capacidades são significativamente reduzidas nos espaços de mensagens privadas, deixando os investigadores e os especialistas na incerteza quanto à disseminação e extensão das narrativas prejudiciais e ao seu potencial impacto na opinião pública [8].

Os espaços públicos das redes sociais permitem aos analistas medir a extensão das campanhas de desinformação e avaliar a amplitude do seu impacto. No entanto, o secretismo das plataformas de mensagens esconde a verdadeira dimensão e demografia da audiência, deixando os investigadores a especular sobre a magnitude do problema. Além disso, identificar as comunidades digitais em que a desinformação prospera tornou-se uma tarefa difícil, uma vez que o acesso a estes espaços é frequentemente limitado.

Nos ambientes mais abertos das redes sociais, era comparativamente mais fácil rastrear a origem das narrativas de desinformação. Com um rasto claro de informação, era possível frequentemente identificar a fonte e as intenções por detrás das campanhas maliciosas. No entanto, as plataformas de mensagens apresentam um desafio de atribuição, uma vez que o conteúdo pode ser reencaminhado e disseminado sem quaisquer marcadores claros da sua origem.

A moderação e a contextualização dos conteúdos desempenham um papel crucial no combate à desinformação, mas as plataformas de mensagens carecem de mecanismos eficazes para estas ações. Ao contrário das plataformas de redes sociais, que têm, mesmo que em escassez, moderadores que podem assinalar, rever e remover conteúdos enganadores, as plataformas de mensagens funcionam sem tais intervenções. Como resultado, a desinformação pode prosperar sem controlo, o que dificulta a redução do seu impacto no discurso público.

Em Portugal, nos primórdios da chegada da Covid-19, foi sobretudo no WhatsApp que circularam as principais narrativas desinformativas sobre o vírus que causava a doença, possíveis curas, mortalidade e origem, entre outros [7]. Investigadores analisaram como as características muito próprias do WhatsApp, tais como a relação de confiança entre membros de grupos e contatos próximos, o efeito de proteção dado por um espaço de comunicação fechado e a facilidade de partilha, podem ser critérios que contribuíram para a distribuição viral de áudios desinformativos, apesar de contraditórios com a comunicação oficial.

No Brasil, nas recentes eleições presidenciais, foi através do WhatsApp e Telegram que foram partilhados a maioria dos conteúdos desinformativos alegando fraudes no processo eleitoral [10]. Nos dias seguintes, a plataforma de mensagens foi a base de mobilização para a ocupação de vários organismos públicos, com apelos ao golpe de Estado contra o presidente eleito, Lula da Silva.

Desinformação pronta a consumir

O principal impacto da IA neste ecossistema pode ser a transformação das campanhas de desinformação num produto autossuficiente e pronto a usar. Significa isto que qualquer utilizador que queira promover ou influenciar uma narrativa desinformativa pode agora produzir e disseminar grandes quantidades de conteúdos de forma rápida e barata. Esta democratização do acesso aos conteúdos pode abrir, inadvertidamente, o caminho para a disseminação de campanhas de influência a uma escala sem precedentes.

O desenvolvimento da IA transformou fundamentalmente a capacidade de produzir conteúdo desinformativo. As ferramentas e os algoritmos alimentados por IA permitem agora a utilizadores com conhecimentos técnicos mínimos gerar material que parece autêntico e fiável. Os algoritmos de geração de texto podem produzir artigos, blogs e publicações nas redes sociais que imitam a linguagem humana, enquanto as imagens e vídeos gerados por IA podem enganar até os olhos mais perspicazes.

Tradicionalmente, a elaboração de campanhas de desinformação exigia muito tempo, esforço e recursos. No entanto, a emergência da IA quebra estas barreiras, permitindo processos rápidos de produção e disseminação. Numa questão de minutos um algoritmo alimentado por IA pode criar ou redirecionar uma rede inteira de perfis falsos, ou gerar um portfolio convincente e heterogéneo de conteúdo multimédia para distribuição. Esta velocidade amplifica significativamente o impacto da desinformação, uma vez que um agente motivado pode inundar o espaço digital com conteúdos desinformativos antes de poderem ser tomadas medidas contrárias.

A nível de recursos, orquestrar uma campanha de desinformação em grande escala costumava exigir um financiamento considerável e capacidades organizacionais. Atualmente, torna-se cada vez mais fácil a qualquer pessoa

com acesso a ferramentas alimentadas por IA lançar, com sucesso, uma campanha de influência, seja isso para eleições para Juntas de Freguesia, concursos de talentos, ou no âmbito da competição entre empresas de uma pequena localidade. É possível antecipar que a democratização e facilitação deste processo vai trazer as operações de desinformação aos combates de opinião de pequenas comunidades e audiências.

Ainda que seja evidente que a IA vai fortalecer muito a nossa capacidade de deteção e análise de desinformação, alguns dos métodos tradicionais de moderação de conteúdos e de verificação de factos poderão, por outro lado, tornar-se menos eficazes face aos conteúdos gerados por IA. Mais do que uma questão de volume, será sobretudo a qualidade que colocará em cheque a nossa capacidade de identificação e mitigação.

Os conteúdos altamente realistas, mas totalmente fabricados, em vídeo, som, imagem e texto irão surgir com cada vez maior frequência. O ainda maior desenvolvimento da tecnologia *Deepfake* aperfeiçoará a criação de vídeos e clips de áudio convincentemente realistas de figuras públicas a dizer ou a fazer coisas que nunca fizeram. Ao explorar a nossa tendência para confiar no que vemos e ouvimos, as campanhas de desinformação podem agora gerar narrativas que serão cada vez mais difíceis de verificar.

Por fim, importa destacar que a IA vai melhorar ainda mais a eficácia dos algoritmos de seleção de conteúdos, permitindo um ainda mais eficiente *micro targeting*, conhecendo provavelmente os nossos vieses, medos e preconceitos até antes de nós próprios [4]. Esses dados serão provavelmente também explorados na exposição a conteúdo desinformativo, mesmo que o algoritmo o faça inadvertidamente, sem consciência que está a expor o utilizador a esse tipo de conteúdo.

Conclusão

As quatro tendências identificadas para o futuro da desinformação – mais pessoal, mais subtil, mais escondida e mais acessível – não são exclusivas, e o carácter paradoxal de algumas delas denota a complexidade deste fenómeno. O seu carácter multidimensional traduz a necessidade de compreensão interdisciplinar e a aposta em diferentes e complementares medidas de

mitigação. O presente capítulo pretende ser um ponto de partida na identificação de algumas características, já identificadas, e para as quais se prevê consolidação. Espera-se que este exercício de antecipação de como a desinformação se adapta às novas realidades do ecossistema digital permita aos leitores, na forma de instituições, investigadores, plataformas e até utilizadores finais, projetar de forma mais eficiente o desenvolvimento de medidas de mitigação.

Bibliografia

1. Borak, M., "The Hunt for Wikipedia's Disinformation Moles", in Wired, 17 de Outubro de 2022. Disponível em https://www.wired.com/story/wikipedia-state-sponsored-disinformation/
2. Cardoso, G., Narciso, I., Moreno, J. & Palma, N., "Online Disinformation During Portugal's 2019 Elections", ISCTE-IUL Media Lab, supported by Democracy Reporting International, November 2019. Disponível em https://democracyreporting.s3.eu-central-1.amazonaws.com/images/2345Portugal-Post-Election-Report-Social-Media-2019.pdf
3. Center for Countering Digital Hate, "The Disinformation Dozen. Why Platforms Must Act on Twelve Leading Online Anti-Vaxxers", 21 de Março de 2021. Disponível em https://counterhate.com/wp-content/uploads/2022/05/210324-The-Disinformation-Dozen.pdf
4. Ienca, M., On Artificial Intelligence and Manipulation, Topoi 42 (2023): 833-842. https://doi.org/10.1007/s11245-023-09940-3
5. Kemp, S. "Digital 2022: Global Overview Report", Datareportal. Disponível em https://datareportal.com/reports/digital-2022-global-overview-report
6. Monteiro, S. B., "Do TikTok para 'O Triângulo': 3 falsidades sobre saúde partilhadas por Alice Santos", in Polígrafo, 2 de Março de 2023. Disponível em https://poligrafo.sapo.pt/saude/artigos/do-tiktok-para-o-triangulo-3-falsidades-sobre-saude-partilhadas-por-alice-santos
7. Moreno, J., Pinto-Martinho, A., Cardoso, G., Narciso, I., Palma, N. e Sepúlveda, R., "Informação e Desinformação sobre o Coronavírus em Portugal – WhatsApp, Facebook e Pesquisas", ISCTE-IUL Media Lab, 2020. Disponível em https://medialab.iscte-iul.pt/informacao-e-desinformacao-sobre-o-coronavirus-em-portugal/
8. Nimmo, B., "The Breakout Scale: Measuring the impact of influence operations", Foreign Policy at Brookings, September 2020. Disponível em https://www.brookings.edu/wp-content/uploads/2020/09/Nimmo_influence_operations_PDF.pdf
9. Powell, A., "Tracing Big Oil's PR war to delay action on climate change", in The Harvard Gazette, 28 de Setembro de 2021. Disponível em https://news.harvard.edu/gazette/story/2021/09/oil-companies-discourage-climate-action-study-says/
10. Santos, J. V., "No ecossistema de desinformação, a centralidade do WhatsApp no adoecimento do sistema democrático. Entrevista especial com João Guilherme Bastos dos Santos", Instituto Humanitas Unisinos, 17 de Janeiro de 2023. Disponível em https://www.ihu.unisinos.br/categorias/159-entrevistas/625597-no-ecossistema-de-desinformacao-a-centralidade-do-whatsapp-no-adoecimento-do-sistema-democratico-entrevista-especial-com-joao-guilherme-bastos-dos-santos
11. Schwartz, M., "A Trail of 'Bread Crumbs', Leading Conspiracy Theorists into the Wilderness", in New York Times, 11 de Setembro de 2018. Disponível em https://www.nytimes.com/2018/09/11/magazine/a-trail-of-bread-crumbs-leading-conspiracy-theorists-into-the-wilderness.html
12. União Europeia, "Tackling online disinformation", 29 de Junho de 2022. Disponível em https://digital-strategy.ec.europa.eu/en/policies/online-disinformation