

Informe de Política para
a Nossa Agenda Comum:

**Integridade da
Informação nas
Plataformas
Digitais**

OUTUBRO DE 2023



**Nações
Unidas**



Introdução

PREFÁCIO

Os desafios atuais apenas podem ser enfrentados por meio de uma cooperação internacional mais forte. A Cúpula do Futuro, que será realizada em 2024, é uma oportunidade de chegar a soluções multilaterais com vistas a um futuro melhor, fortalecendo a governança global para as gerações presentes e futuras (Resolução da Assembleia Geral nº 76/307). Na minha qualidade de secretário-geral, fui convidado a contribuir para os preparativos da Cúpula, fazendo recomendações orientadas para a ação, com base nas propostas contidas em meu relatório intitulado “Nossa Agenda Comum” (A/75/982), que foi uma resposta à declaração da comemoração do septuagésimo quinto aniversário das Nações Unidas (Resolução da Assembleia Geral nº 75/1). O presente Informe é uma dessas contribuições. Ele compila as ideias que foram primeiramente propostas na Nossa Agenda Comum, que levou em consideração a orientação dos Estados-membros e as consultas intergovernamentais e multissetoriais realizadas no decorrer de mais de um ano. Elas foram fundamentadas em propósitos e princípios da Carta das Nações Unidas, da Declaração Universal dos Direitos Humanos e outros instrumentos internacionais.

OBJETIVO DESTA INFORME

O presente Informe está focado em como as ameaças à integridade da informação estão tendo um impacto no progresso em questões globais, nacionais e locais. Na Nossa Agenda Comum, eu propus um consenso empírico, que se baseia em torno de fatos, ciência e conhecimento. Para tanto, o presente Informe descreve os princípios potenciais para um Código de Conduta que auxiliará a orientar os Estados-membros, as plataformas digitais e outros atores interessados nos esforços para tornar o espaço digital mais inclusivo e seguro para todas as pessoas, ao passo que defende vigorosamente o direito à liberdade de opinião e expressão e o direito de acesso à informação. O Código de Conduta para a Integridade da Informação em Plataformas Digitais está sendo desenvolvido no âmbito dos preparativos para a Cúpula do Futuro. A minha esperança é que a Cúpula forneça um padrão de excelência para orientar ações que visam fortalecer a integridade da informação.

As plataformas digitais são ferramentas cruciais que transformaram as interações sociais, culturais e políticas em todos os lugares. Em todo o mundo, elas conectam cidadãos globais interessados em questões importantes. As plataformas auxiliam as Nações Unidas a informar e engajar diretamente as pessoas em nossos esforços pela paz, dignidade e igualdade em um planeta saudável. Elas deram esperança às pessoas em tempos de crise e luta, amplificaram vozes que antes não eram ouvidas e deram vida a movimentos globais.

No entanto, essas mesmas plataformas também expuseram um lado mais sombrio do ecossistema digital. Elas permitiram a rápida disseminação de mentiras e do discurso de ódio, causando danos reais em escala global. O otimismo sobre o potencial das redes sociais para conectar e engajar as pessoas foi diminuindo à medida que a desinformação e o discurso de ódio foram se movendo das margens ao centro do espaço digital. O perigo não pode ser subestimado. O discurso de ódio e a desinformação possibilitados pelas redes sociais podem levar à violência e à morte.¹ A capacidade de disseminar desinformação em

larga escala para minar fatos cientificamente estabelecidos representa um risco existencial para a humanidade (A/75/982, parágrafo 26) e coloca em perigo as instituições democráticas e os direitos humanos fundamentais. Esses riscos se intensificaram ainda mais devido aos rápidos avanços da tecnologia, como a inteligência artificial generativa. Em todo o mundo, as Nações Unidas estão monitorando como a desinformação e o discurso de ódio podem ameaçar o progresso dos Objetivos de Desenvolvimento Sustentável (ODS), deixando claro que mudanças são imprescindíveis.

O presente documento foi preparado pelo secretário-geral das Nações Unidas, António Guterres, como parte da [série de onze Informes de Política para a Nossa Agenda Comum](#). Este Informe foi publicado originalmente em inglês em junho de 2023, marcando o segundo [Dia Internacional de Combate ao Discurso de Ódio](#).

A convite dos Estados-membros, o secretário-geral preparou uma série de Informes para fornecer mais detalhes sobre determinadas propostas contidas no documento [Nossa Agenda Comum](#) e para apoiá-los em suas deliberações enquanto se preparam para a Cúpula do Futuro, que será realizada em setembro de 2024. Os Informes incluem análises do impacto das propostas na implementação dos Objetivos de Desenvolvimento Sustentável (ODS) e são fundamentados na Carta das Nações Unidas e na Declaração Universal dos Direitos Humanos, tendo a igualdade de gênero como tema transversal.

Tradução para o português: Jéssica Monteiro Silva, Programa de Voluntários das Nações Unidas (UNV)
Revisão e publicação no Brasil: Centro de Informação das Nações Unidas para o Brasil (UNIC Rio)

NA PRÁTICA, A DIFERENÇA ENTRE
INFORMAÇÃO FALSA E DESINFORMAÇÃO

PODE SER SUTIL E

DIFÍCIL DE

DETERMINAR

O que é integridade da informação?

A integridade da informação refere-se à precisão, consistência e confiabilidade da informação. Ela é ameaçada pela desinformação, pela informação falsa e discurso de ódio. Embora não haja definições universalmente aceitas desses termos, as entidades das Nações Unidas desenvolveram definições de trabalho dos mesmos.

A relatora especial sobre a Promoção e Proteção do Direito à Liberdade de Opinião e Expressão refere-se à desinformação como “informações falsas que são divulgadas intencionalmente para causar sérios danos sociais”.² A desinformação é descrita pela Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO) como conteúdo falso ou enganoso que pode causar danos específicos, independentemente de motivações, consciência ou comportamentos.³

Para os propósitos do presente Informe, a diferença entre a informação falsa e a desinformação está na intenção:

- A desinformação é uma informação que não é apenas imprecisa, mas também tem a intenção de enganar e é espalhada para causar danos. A desinformação pode ser disseminada por atores estatais ou não estatais em múltiplos contextos, inclusive durante conflitos armados, e pode afetar todas as áreas do desenvolvimento, desde paz e segurança até direitos humanos, saúde pública, ajuda humanitária e ação climática.
- A informação falsa se refere à disseminação não intencional de informações imprecisas, compartilhadas de boa fé por aqueles que não sabem que estão transmitindo falsidades.

A desinformação pode estar enraizada em informações falsas, como mentiras e narrativas

enganosas transmitidas involuntariamente pelas pessoas ao longo do tempo que acabam sendo utilizadas deliberadamente como armas pelo discurso político.⁴

Na prática, a distinção entre a informação falsa e a desinformação pode ser difícil de determinar.

Discurso de ódio, de acordo com a definição de trabalho da Estratégia e Plano de Ação das Nações Unidas sobre o Discurso de Ódio, é “qualquer tipo de comunicação oral, escrita ou comportamento, que ataca ou usa linguagem pejorativa ou discriminatória com referência a uma pessoa ou grupo com base em quem eles são, ou seja, com base em sua religião, etnia, nacionalidade, raça, cor, descendência, gênero ou outro fator identitário”.⁵

A informação falsa, a desinformação e o discurso de ódio são fenômenos relacionados, mas distintos, com certas áreas de sobreposição e diferenças na forma como podem ser identificados, mitigados e abordados. Todos os três poluem o ecossistema da informação e ameaçam o progresso humano.⁶

As ameaças à integridade da informação não são novas. Mentiras e ódio têm sido espalhados há muito tempo para ganhos políticos ou financeiros.

No entanto, na era digital, essas operações podem ser realizadas em uma escala que antes era impensável. Poderosas ferramentas de comunicação agora podem espalhar conteúdo instantaneamente por todo o mundo, criando um problema tão generalizado que as próprias plataformas digitais às vezes são incapazes de compreender sua extensão total. A falta de definições acordadas em nível intergovernamental não deve resultar em inércia. Devemos fazer tudo o que pudermos para conter os danos que eles causam.

Integridade da informação e plataformas digitais

As plataformas digitais devem ser atores integrais no esforço de manter a integridade da informação. Embora certas mídias tradicionais também possam ser fontes de desinformação, a velocidade, o volume e o efeito viral de sua disseminação por meio de canais digitais exigem uma resposta urgente e personalizada. Para efeitos do presente Informe, o termo “plataforma digital” refere-se a um serviço digital que facilita as interações entre dois ou mais usuários, abrangendo uma ampla gama de atividades, desde redes sociais e mecanismos de busca até aplicativos de mensagens. Normalmente, elas coletam dados sobre seus usuários e suas interações.⁷

A informação falsa e a desinformação são criadas por uma ampla gama de atores, com várias motivações, que em geral, conseguem permanecer anônimos.

Campanhas de desinformação coordenadas por atores estatais e não estatais têm explorado sistemas digitais falhos para promover narrativas nocivas, com sérias repercussões.

Muitos Estados lançaram iniciativas para regulamentar as plataformas digitais, com pelo menos 70 dessas leis adotadas ou consideradas nos últimos quatro anos.⁸ Em sua essência, as abordagens legislativas geralmente envolvem um escopo restrito de medidas e ações para definir e remover conteúdo nocivo. Ao se concentrar na remoção de conteúdo prejudicial, alguns Estados introduziram uma legislação deficiente e ampla demais que, na verdade, silenciou o “discurso protegido”, permitido pelo direito internacional.

Outras respostas, como bloqueios gerais da Internet e proibições de plataformas, podem carecer de base legal e infringir os direitos humanos. Muitos Estados e figuras políticas alegaram preocupações

com a integridade da informação como pretexto para restringir o acesso à informação, desacreditar e restringir reportagens e atingir jornalistas e opositores.⁹ Os atores estatais também pressionaram as plataformas a fazerem suas vontades sob o pretexto de combater a informação falsa e a desinformação.¹⁰

Especialistas em liberdade de expressão enfatizaram que os Estados têm um dever especial neste contexto e “não devem criar, patrocinar, encorajar ou promover informações falsas” (A/77/287, parágrafo 45).

Os riscos inerentes à regulamentação da expressão exigem uma abordagem cuidadosamente adaptada que atenda aos requisitos de legalidade, necessidade e proporcionalidade da lei de direitos humanos, mesmo quando houver um propósito legítimo de interesse público (ibid., parágrafo. 42).

A desinformação também é um grande negócio. Tanto as empresas de relações públicas “obscuras” quanto as tradicionais, contratadas por Estados, figuras políticas e pelo setor privado, são as principais fontes de conteúdo falso e enganoso.¹¹ Uma tática, entre outras, tem sido a de publicar conteúdo em versões clonadas falsas de sites de notícias para fazer com que os artigos pareçam ser de fontes legítimas.¹² Esse negócio obscuro é extremamente difícil de rastrear e pesquisar, de modo que a verdadeira escala do problema é desconhecida. Indivíduos também espalham informações falsas para vender produtos ou serviços com fins lucrativos, muitas vezes visando grupos vulneráveis em tempos de crise e/ou insegurança.

Abordagem dominante dos atuais modelos de negócios da maioria das plataformas digitais gira em torno da “economia da atenção”. Algoritmos são projetados para priorizar o conteúdo que mantém a atenção dos usuários, maximizando assim o engajamento e o faturamento publicitário. O conteúdo impreciso e discurso de ódio concebidos para polarizar os usuários e gerar emoções fortes costuma ser o que gera mais engajamento. Como resultado, algoritmos são conhecidos por recompensar e ampliar a informação falsa, a desinformação e o discurso de ódio.¹³

Enfrentando um declínio na receita de publicidade, as plataformas digitais estão buscando caminhos alternativos para lucrar além da economia da atenção. Por exemplo, planos de verificação pagos, de acordo com os quais as contas podem comprar um selo de aprovação usado anteriormente para denotar autenticidade, levantaram sérias preocupações em relação à integridade da informação, devido ao potencial de abuso por agentes de desinformação.¹⁴

2 BILHÕES

DE PESSOAS USAM AS REDES SOCIAIS

Fonte: Kepios, Digital 2023: Global Overview Report (2023)

TWITTER
0.6B

FACEBOOK MESSENGER
0.9B



FACEBOOK

3.0B

SNAPCHAT
0.6B

DOUYIN
0.7B

KUAISHOU
0.6B

TIKTOK
1.1B

WHATSAPP
2.0B

PINTEREST
0.4B

SINO WEIBO
0.6B

WECHAT
1.3B

TELEGRAM
0.7B

YOUTUBE
2.5B

QQ
0.6B

INSTAGRAM
2.0B

Qual é o quadro jurídico internacional relevante?

A promoção da integridade da informação deve estar totalmente fundamentada nas normas e padrões internacionais pertinentes, incluindo a lei de direitos humanos e os princípios de soberania e não intervenção em assuntos domésticos. Em agosto de 2022, enviei à Assembleia Geral um relatório intitulado “Combatendo a desinformação para a promoção e proteção dos direitos humanos e liberdades fundamentais”.¹⁵ No relatório, expliquei a lei internacional de direitos humanos que se aplica à desinformação, incluindo a Declaração Universal dos Direitos Humanos e o Pacto Internacional sobre Direitos Civis e Políticos. De acordo com esses instrumentos jurídicos internacionais, todos têm direito à liberdade de expressão.¹⁶

O Artigo 19 da Declaração Universal dos Direitos Humanos e o Artigo 19 (2) do Pacto protegem o direito à liberdade de expressão, incluindo a liberdade de procurar, receber e difundir informações e ideias de todos os tipos, independentemente de fronteiras e por qualquer meio de comunicação. O direito humano à liberdade de expressão não se limita às informações recebidas favoravelmente (A/77/287, parágrafo 13). Ligada à liberdade de expressão, a liberdade de informação é em si um direito. A Assembleia Geral declarou: “A liberdade de informação é um direito humano fundamental e é o marco de todas as liberdades às quais as Nações Unidas consagraram” (ibid., para. 14). A liberdade de expressão e o acesso à informação podem estar sujeitos a certas restrições que atendem a critérios específicos estabelecidos no artigo 19 (3) do Pacto.¹⁷ Os Estados não podem incluir motivos adicionais ou restringir a liberdade de expressão além do que é permitido pelo direito internacional.

O Plano de Ação de Rabat sobre a proibição da defesa do ódio nacionalista, racial ou religioso que constitua incitação à hostilidade, discriminação ou violência, adotado em 2012, fornece orientação prática jurídica e política aos Estados sobre a melhor forma de implementar o Artigo 20 (2) do Pacto e o Artigo 4 da Convenção Internacional sobre a Eliminação de Todas as Formas de Discriminação Racial, que proíbe certas formas de discursos de ódio. O Plano de Ação de Rabat já foi utilizado pelos Estados-Membros em diferentes contextos.¹⁸

O discurso de ódio tem sido um precursor pra crimes de atrocidade, incluindo o genocídio. A Convenção de 1948 sobre a Prevenção e Punição do Crime de Genocídio proíbe “a incitação direta e pública à prática de genocídio”.

Na resolução 76/227, adotada em 2021, a Assembleia Geral enfatizou que todas as formas de desinformação podem impactar negativamente o gozo dos direitos humanos e das liberdades fundamentais, bem como o alcance dos Objetivos de Desenvolvimento Sustentável. Similarmente, na resolução 49/21, adotada em 2022, o Conselho de Direitos Humanos afirmou que a desinformação pode afetar negativamente o gozo e a efetivação de todos os direitos humanos.

O DISCURSO DE ÓDIO

TEM SIDO UM PRECURSOR PARA

CRIMES DE ATROCIDADE

INCLUINDO O

GENOCÍDIO

Qual é o dano causado pela informação falsa, pela desinformação on-line e pelo discurso de ódio?

A informação falsa, a desinformação e o discurso de ódio on-line representam sérias preocupações para pessoas em todo o mundo. Um estudo realizado com um grupo de entrevistados em 142 países mostrou que 58,5% dos usuários regulares de Internet e mídias sociais em todo o mundo estão preocupados em encontrar informações falsas on-line, sendo que os jovens e pessoas de baixa renda se sentem significativamente mais vulneráveis.¹⁹ Os jovens de hoje são nativos digitais, com maior probabilidade de se conectarem on-line do que o resto da população, tornando-os a geração mais conectada digitalmente da história.²⁰ Em todo o mundo, uma criança fica on-line pela primeira vez a cada meio segundo, ficando sob o risco de exposição a discursos de ódio e danos on-line, em alguns casos afetando a saúde mental delas.²¹

Os impactos das informações falsas, da desinformação e do discurso de ódio on-line podem ser vistos em todo o mundo, inclusive nas áreas da saúde, ação climática, democracia e eleições, igualdade de gênero, segurança e resposta humanitária. A poluição da informação foi identificada como uma preocupação significativa por 75% dos escritórios nacionais do Programa das Nações Unidas para o Desenvolvimento (PNUD) em uma pesquisa realizada em 2021.

ESTUDO REALIZADO EM

142 PAÍSES MOSTROU QUE

58.5%

DOS USUÁRIOS REGULARES DE INTERNET E MÍDIAS SOCIAIS AO REDOR DO MUNDO ESTÃO PREOCUPADOS EM ENCONTRAR

DESINFORMAÇÃO ON-LINE

Isso tem graves implicações na confiança, segurança, democracia e desenvolvimento sustentável, conforme constatado em uma recente revisão encomendada pela UNESCO de mais de 800 documentos acadêmicos, da sociedade civil, jornalísticos e corporativos.²²

A informação falsa e a desinformação podem ser perigosas e potencialmente mortais, especialmente em tempos de crise, emergência ou conflito. Durante a pandemia da COVID-19, um dilúvio de informação falsa sobre o vírus, medidas de saúde pública e vacinas começou a circular on-line.²³ Certos atores exploraram o mal-entendido visando seus próprios interesses, com ativistas anti-vacina levando usuários a sites que vendiam falsas curas ou medidas preventivas enganosas.²⁴ Muitas das vítimas da COVID-19 se recusaram a ser vacinadas ou a tomar precauções básicas de saúde depois de serem expostas a informações falsas on-line.²⁵

A informação falsa também pode ser mortal em contextos sociais e políticos já voláteis. Em um relatório de 2022, o relator especial sobre a Promoção e Proteção do Direito à Liberdade de Opinião e Expressão examinou o impacto da utilização da informação como arma para semear desordem, alimentar o ódio, incitar a violência e prolongar um conflito.²⁶ Outro relatório divulgado em 2022, revelou que a informação falsa pode “envolver fanatismo e discurso de ódio direcionado a minorias, mulheres e todos os chamados ‘outros’, representando ameaças não apenas às pessoas diretamente visadas, mas também à inclusão e coesão social. Isso pode ampliar tensões e divisões

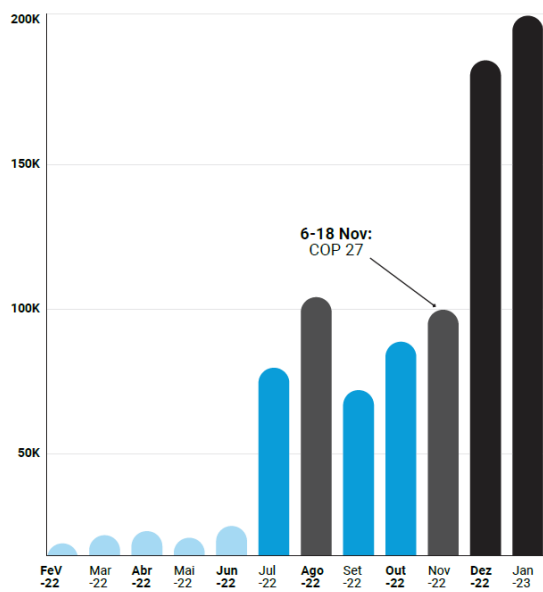
em tempos de emergência, crise, momentos políticos chave ou durante conflitos armados”.²⁷

Alguns dos piores impactos dos danos on-line ocorrem em lugares e contextos negligenciados pelas plataformas digitais, países em meio a conflitos, ou com contextos voláteis que muitas vezes são mercados menos lucrativos, não direcionam recursos suficientes para moderação de conteúdo ou assistência ao usuário. Embora a mídia tradicional continue sendo uma importante fonte de notícias para a maioria das pessoas em áreas de conflito, o ódio espalhado nas plataformas digitais também tem desencadeado e impulsionado a violência.²⁸ Algumas plataformas digitais foram criticadas por suas atuações em conflitos, incluindo a guerra em curso na Ucrânia.²⁹

Da mesma forma, informações errôneas e desinformação sobre a emergência climática estão atrasando ações urgentemente necessárias para garantir um futuro habitável no planeta Terra. A informação falsa e a desinformação sobre o clima podem ser entendidas como conteúdo falso ou enganoso que mina a base cientificamente aceita sobre a existência da mudança climática induzida pela humanidade, suas causas e impactos.

FIGURA I

NÚMERO MENSAL DE USOS DA #CLIMATESCAM (#FARSADOCLIMA) ON NO TWITTER



Fonte: Departamento de Comunicação Global (DGC) da ONU, usando dados do Talkwalker.

Campanhas coordenadas estão tentando negar, minimizar ou desviar a atenção do consenso científico do Painel Intergovernamental sobre Mudanças Climáticas (IPCC) e inviabilizar ações urgentes para atingir as metas do Acordo de Paris de 2015. Uma minoria, porém expressiva, de negacionistas da ciência do clima³⁰ continua rejeitando a posição de consenso e possui uma presença descomunal em algumas plataformas digitais. Por exemplo, em 2022, simulações aleatórias de organizações da sociedade civil revelaram que o algoritmo do Facebook estava recomendando conteúdo negacionista do clima em detrimento da ciência climática.³¹ No Twitter, os usos da hashtag #climatescam (#farsadoclima) dispararam de menos de 2.700 por mês no primeiro semestre de 2022 para 80.000 em julho e 199.000 em janeiro de 2023. A frase também foi destaque na plataforma, estando entre os principais resultados na busca por “clima”.³² Em fevereiro de 2022, o Painel Intergovernamental sobre Mudanças Climáticas designou pela primeira vez a “informação falsa climática”, afirmando que o “enfraquecimento deliberado da ciência” estava contribuindo para “percepções errôneas sobre o consenso científico e incerteza, além da desconsideração do risco e da urgência”.³³

Algumas empresas de combustíveis fósseis geralmente implementam uma estratégia de “lavagem verde” (‘greenwashing’), levando o público a acreditar que uma empresa ou entidade está fazendo mais para proteger o meio ambiente do que o prejudicando. As companhias não atuam sozinhas. Esforços para confundir o público e desviar a atenção da responsabilidade da indústria de combustíveis fósseis são possibilitados e apoiados por agências de publicidade e relações públicas, empresas de tecnologias de comunicação, agências de notícias e plataformas digitais.³⁴ As empresas de publicidade e relações públicas que criam conteúdo de lavagem verde e terceiros que o distribuem estão ganhando bilhões com esses esforços para proteger a indústria de combustíveis fósseis do escrutínio público e de sua responsabilidade. Empresas de relações públicas realizaram centenas de campanhas para empresas de carvão, petróleo e gás.³⁵

70%

DOS MEMBROS DAS FORÇAS DE PAZ DISSERAM QUE A

INFORMAÇÃO FALSA E A DESINFORMAÇÃO

ESTÃO TENDO UM IMPACTO

SEVERO, CRÍTICO

OU MODERADO NO SEU TRABALHO

A informação falsa e a desinformação estão tendo um impacto profundo na democracia, enfraquecendo a confiança nas instituições democráticas e na mídia independente, e enfraquecendo a participação cidadã em assuntos políticos e públicos. Ao longo do ciclo eleitoral, a exposição a informações falsas e enganosas pode privar os eleitores da chance de fazer escolhas informadas. A disseminação de informação falsa pode minar a confiança do público nas instituições eleitorais e no próprio processo eleitoral – como no registro eleitoral, na votação e nos resultados – e potencialmente resultar em apatia do eleitor ou rejeição de resultados eleitorais credíveis. Estados e líderes políticos provaram ser fontes potentes de informação falsa, espalhando deliberadamente e estrategicamente mentiras para manter ou assegurar o poder, ou minar processos democráticos em outros países.³⁶

Grupos marginalizados e vulneráveis também são alvos frequentes de informação falsa e do discurso de ódio, resultando em sua exclusão social, econômica e política. Candidatas, eleitoras, funcionárias eleitorais, jornalistas e representantes da sociedade civil são alvos de informação falsa de gênero on-line.³⁷ Esses ataques prejudicam a participação política e enfraquecem as instituições democráticas e os direitos humanos, incluindo a liberdade de expressão e acesso à informação desses grupos.³⁸ Esta deve ser uma prioridade cada vez mais urgente para a comunidade internacional, até porque mais de 2 bilhões de eleitores devem ir às urnas em todo o mundo em 2024.

A falta de informação e a desinformação também se multiplicam entre as plataformas digitais e a mídia tradicional, tornando-se ainda mais complexas de rastrear e resolver se não forem detectados

na fonte. A desinformação pode ser uma tática deliberada de meios de comunicação influenciados ideologicamente, cooptados por interesses políticos e corporativos.³⁹ Ao mesmo tempo, a ascensão das plataformas digitais precipitou um declínio dramático da mídia confiável e independente. O público das notícias e as receitas publicitárias migraram em massa para as plataformas da Internet – uma tendência exacerbada pela pandemia da COVID-19. “extinção da mídia” ou “desertos de notícias”, em que as comunidades perdem fontes confiáveis de notícias locais, podem ser vistas em algumas regiões ou países,⁴⁰ contribuindo para a poluição do ecossistema de informação. A “lavagem de notícias” – em que o conteúdo patrocinado é disfarçado para se parecer com notícias relatadas – é muitas vezes sinalizada inadequadamente quando postada em plataformas digitais, conferindo-lhe uma aparência de legitimidade. Uma vez captada por outros meios de comunicação, citada pela classe política, ou amplamente compartilhada em plataformas digitais, a origem da informação torna-se cada vez mais obscura e os consumidores de notícias ficam incapazes de distingui-la de um fato real.

A informação falsa também está tendo um impacto direto no trabalho das Nações Unidas. Os Coordenadores Residentes, enviados, mediadores e Forças de Paz levantaram preocupações sobre o efeito da desinformação na segurança operacional, eficácia e capacidade de entrega da Organização. Em uma pesquisa de 2022, 70 por cento dos membros entrevistados em Forças de Paz das Nações Unidas disseram que a desinformação estava tendo um impacto grave e crítico ou moderado em seu trabalho, enquanto 75 por cento disseram que estavam tendo um impacto em sua segurança e proteção. A informação falsa também pode ser usada para atingir trabalhos humanitários e dificultar as operações de salvamento em áreas de conflito.

FIGURA II

INTEGRIDADE DA INFORMAÇÃO E OBJETIVOS DE DESENVOLVIMENTO SUSTENTÁVEL

Conforme demonstrado, as ameaças à integridade da informação podem ter um impacto negativo na implementação de todos os 17 Objetivos de Desenvolvimento Sustentável.



A informação falsa e a desinformação continuam a ter implicações nos esforços de erradicação da pobreza e na economia. As dificuldades econômicas também podem alimentar a disseminação de mentiras polarizadoras e odiosas, inclusive sobre grupos marginalizados. A crise do custo de vida tem sido um terreno particularmente fértil para disseminação de informação enganosa, culpando falsamente as fontes de energia renovável pelo aumento dos custos de energia ou perda de empregos, por exemplo.



A informação falsa, a desinformação e o discurso de ódio disseminados on-line estão polarizando as sociedades, mirando comunidades já marginalizadas e vulneráveis, resultando numa maior exclusão social, econômica e política.



Ameaças à integridade da informação podem agravar a fome, inclusive por exacerbar conflitos, inibir ações climáticas e de resposta à desastres, à pobreza e à desigualdade. A informação falsa pode desviar a atenção dos desafios criados pelos conflitos à segurança alimentar global.



A informação falsa, a desinformação e o discurso de ódio e as respostas excessivamente vagas a esses fenômenos podem ter um impacto negativo na inovação, inclusive limitando o potencial de grupos marginalizados e tornando os espaços digitais menos igualitários e inclusivos.



Durante a pandemia da COVID-19, uma "infodemia" relacionada à falta de informação e à desinformação prejudicou as medidas de saúde pública e as campanhas de vacinação. A poluição da informação ameaça a saúde e o bem-estar das crianças expostas a conteúdos nocivos.



Os esforços para tornar as cidades e as comunidades mais sustentáveis podem ser prejudicados pela desinformação que nega ou desvia a atenção dos impactos da atividade humana no meio ambiente.



A informação falsa, a desinformação e o discurso de ódio podem ter um impacto adverso no acesso à educação de qualidade, em particular para grupos marginalizados, incluindo mulheres e meninas. O acesso à informação e à alfabetização em mídias digitais pode aumentar a resiliência social e desempenhar um papel fundamental na limitação do impacto social dos danos da poluição da informação on-line.



A desinformação climática e a inércia que ela incentiva estão minando os esforços para tomar medidas urgentes para enfrentar a crise climática, inclusive impedindo a mudança crucial de combustíveis fósseis poluentes para fontes de energia renovável e investimentos urgentes em adaptação e resiliência climática.



O discurso de ódio e a informação falsa baseados em gênero buscam subjugar sistematicamente as mulheres, silenciando-as e expulsando-as da esfera pública. Isso pode ter consequências devastadoras, desde suprimir as vozes das mulheres e alimentar a autocensura até causar danos profissionais e à reputação, além de incitar a violência física.



A informação falsa e a desinformação podem ter um impacto negativo nos esforços para proteger os oceanos e mares e usar de forma sustentável os recursos marinhos.



Dois bilhões de pessoas vivem sem serviços de água potável administrados com segurança. A informação falsa e a desinformação sobre a segurança da água potável e do saneamento podem ter consequências perigosas para a saúde.



Ativistas ambientais que trabalham para proteger a vida na terra foram alvo de campanhas de ódio, da informação falsa e da desinformação on-line, com consequências na vida real. A desinformação está também prejudicando os esforços de ação climática.



A informação falsa e a desinformação sobre o clima, em grande parte semeada pela indústria de combustíveis fósseis, está prejudicando a transição urgente para formas mais limpas de produção de energia, estreitando a janela de oportunidade para oferecer um futuro sustentável para todas as pessoas em um planeta habitável.



A desinformação e o discurso de ódio têm sido usados para influenciar eleições e narrativas públicas e semear a confusão. Eles têm sido usados para minar os adversários, frustrar os esforços de pacificação, incitar a violência, prolongar conflitos e prejudicar a confiança no Estado de Direito. Os esforços para promover sociedades pacíficas e inclusivas e o papel das Nações Unidas em apoiar a paz e a segurança foram seriamente afetados como resultado.



Pesquisas já demonstram impactos prejudiciais da falta de informação, da desinformação e do discurso de ódio no crescimento econômico.



A informação falsa, a desinformação e o discurso de ódio podem impedir parcerias significativas para alcançar os Objetivos de Desenvolvimento Sustentável, reduzir a alocação de recursos necessários e enfraquecer os esforços para não deixar ninguém para trás.



A informação falsa, a desinformação e o discurso de ódio e as respostas excessivamente vagas a esses fenômenos podem ter um impacto negativo na inovação, inclusive limitando o potencial de grupos marginalizados e tornando os espaços digitais menos igualitários e inclusivos.

a World Health Organization and United Nations Children's Fund (UNICEF), Progress on Household Drinking Water, Sanitation and Hygiene 2000–2020: Five Years into the SDGs (Genebra, 2021).
b Veja Roberto Cavazos e CHEQ, "The economic cost of bad actors on the Internet: fake news, 2019"; e London Economics, "The cost of lies: assessing the human and financial impact of COVID-19 related online misinformation on the UK", Dezembro, 2020.
c Global Witness, Last Line of Defence: The Industries Causing the Climate Crisis and Attacks against Land and Environmental Defenders (2021).
d A/77/288.

Como fortalecer a integridade da informação?

A informação falsa, a desinformação e o discurso de ódio não existem no vácuo. Elas se espalham quando as pessoas se sentem excluídas e/ou não ouvidas, quando são confrontadas com os impactos da disparidade econômica, e quando se sentem politicamente desencantadas. As respostas à poluição da informação devem abordar esses desafios do mundo real. Esforços para alcançar os Objetivos de Desenvolvimento Sustentável são fundamentais para construir um mundo em que a confiança possa ser restaurada.

Ao elaborar respostas, é importante não perder de vista o enorme valor que as plataformas digitais trazem para o mundo. As plataformas revolucionaram a comunicação de massa em tempo real, permitindo a disseminação de informações que salvam vidas durante desastres naturais e pandemias. Elas ajudaram a mobilizar apoio para os objetivos pelos quais as Nações Unidas lutam, muitas vezes provando ser forças positivas para a inclusão e a participação na vida pública. Elas conectaram comunidades geograficamente distantes de pessoas excluídas, incluindo aquelas que sofrem de condições de saúde raras, e uma vasta gama de ativistas que trabalham para tornar o mundo um lugar melhor.

RESPOSTAS REGULATÓRIAS

A questão de saber se as plataformas digitais podem e devem ser responsabilizadas legalmente pelo conteúdo que hospedam tem sido objeto de um longo debate. Em alguns contextos, as leis existentes baseadas em difamação, assédio digital ('cyberbullying') foram usadas de forma eficaz para combater ameaças à integridade da informação sem impor novas restrições à liberdade de expressão ([A/77/287](#), parágrafo 44).

Além disso, recentes esforços legislativos foram feitos para abordar a questão nos níveis regional e nacional. Isso inclui a estrutura adotada pela União Europeia em 2022, que compreende a Lei de Serviços Digitais, a iniciativa sobre transparência e direcionamento de publicidade política e o Código de Conduta sobre Desinformação. A Lei dos Serviços Digitais estabelece novas regras para usuários, plataformas digitais e empresas que operam on-line na União Europeia. As medidas miram o conteúdo e os bens e serviços on-line ilegais, além de fornecerem um mecanismo para que os usuários denunciem o conteúdo ilegal e contestem as decisões de moderação. Elas exigem que as plataformas digitais melhorem a transparência, especialmente no uso e na natureza dos algoritmos de recomendação, e que plataformas maiores forneçam aos pesquisadores acesso a dados.

O Código de Conduta sobre Desinformação da União Europeia estabelece princípios e compromissos para plataformas on-line e para o setor de publicidade com a finalidade de combater a disseminação de desinformação on-line, os quais seus signatários concordaram em implementar.⁴¹ Esse Código inclui compromissos voluntários que objetivam a desmonetização da informação falsa, tanto impedindo a disseminação de anúncios contendo informação falsa, quanto evitando a veiculação de anúncios junto com conteúdo contendo informação falsa. Os signatários também concordaram em rotular a publicidade política de forma mais clara, juntamente com detalhes do patrocinador, gastos com publicidade e período de exibição, além de criar bancos de dados para pesquisas sobre propagandas políticas. Além disso, eles se comprometeram a compartilhar informações

sobre comportamentos manipulativos e maliciosos usados para espalhar desinformação (como contas falsas, ampliação por bots, usurpação de identidade e deepfakes maliciosas) detectados em suas plataformas, ademais de atualizar e implementar regularmente políticas para combatê-los. Outros compromissos estão focados em treinar os usuários a reconhecer, entender e sinalizar a desinformação, fortalecer a colaboração com verificadores de fatos e fornecer aos pesquisadores um melhor acesso aos dados. O teste real desses novos mecanismos será sua implementação.

RESPOSTAS DAS PLATAFORMAS DIGITAIS

Um dos principais objetivos do Código de Conduta sobre Desinformação é melhorar a transparência das plataformas digitais. Em fevereiro de 2023, os signatários do Código de Conduta publicaram seus relatórios iniciais sobre como estão implementando os compromissos. Os relatórios apresentaram informações sobre até que ponto os rendimentos publicitários foram impedidos de circular informação para atores de desinformação e outros comportamentos manipuladores detectados, incluindo um esforço coordenado em larga escala para manipular a opinião pública sobre a guerra na Ucrânia em vários países europeus.⁴²

As plataformas digitais são muito diversas em termos de tamanho, função e estrutura e têm buscado uma ampla gama de respostas para combater os danos. Várias das maiores plataformas se comprometeram publicamente a defender os Princípios Orientadores sobre Empresas e Direitos Humanos das Nações Unidas,⁴³ porém lacunas na política, transparência e implementação persistem. Algumas plataformas não seguem seus próprios padrões e, em graus variados, permitem e ampliam as mentiras e o ódio.⁴⁴ Algoritmos criados para promover o modelo lucrativo das plataformas são projetados para maximizar deliberadamente o engajamento

e monopolizar a atenção, tendendo a empurrar os usuários para conteúdos polarizadores ou provocativos.

A maioria das plataformas digitais tem algum tipo de sistema de autorregulação, moderação ou mecanismos de supervisão em vigor, mas a transparência em torno da política e das práticas de remoção de conteúdo continua sendo um desafio.⁴⁵ O investimento nesses mecanismos em todas as regiões e idiomas é extremamente irregular e amplamente concentrado no Norte Global, assim como a fiscalização de suas próprias regras pelas plataformas. Uma pesquisa recente revelou que a tradução de ferramentas de moderação e mecanismos de supervisão para os idiomas locais é incompleta nas plataformas.⁴⁶ Ao mesmo tempo, a moderação é muitas vezes terceirizada e lamentavelmente carente de recursos em outros idiomas além do inglês.⁴⁷

O testemunho dos moderadores levantou questões preocupantes relacionadas a maus-tratos, padrões de trabalho e traumas secundários.⁴⁸ Os moderadores relatam estar constantemente expostos a conteúdo violento e perturbador, e ter de determinar se uma postagem denunciada viola a política da empresa em questão de segundos. Os sistemas automatizados de moderação de conteúdo podem desempenhar um papel essencial, mas estão expostos a possíveis viesamentos com base nos dados e estruturas usadas para treiná-los. Eles também têm altas taxas de erro em inglês e são ainda piores em outros idiomas. Várias plataformas digitais empregam equipes de confiança e segurança, que possuem formação em direitos humanos e na integridade da informação, mas esses especialistas geralmente não são incluídos nos estágios iniciais do desenvolvimento dos produtos e costumam ser os primeiros a terem os empregos cortados durante medidas de economia e contenção de gastos.

ACESSO A DADOS

O acesso a dados para pesquisadores também é uma prioridade urgente em escala global. A pesquisa e os recursos existentes permanecem fortemente direcionados para os Estados Unidos e para a Europa. Com notáveis exceções, incluindo os relatórios sobre as Forças de Manutenção da Paz das Nações Unidas na África e a missão internacional independente de apuração de fatos em Mianmar,⁴⁹ e algumas reportagens e coberturas investigativas de jornalistas,⁵⁰ um número limitado de pesquisas foram publicadas sobre o impacto no resto do mundo. Isso ocorre em parte porque os pesquisadores não têm acesso às plataformas e aos seus dados. As ferramentas necessárias para pesquisas a partir dos dados limitados fornecidos pelas plataformas também tendem a ser projetadas com foco nos resultados em termos de publicidade, e são excessivamente caras. Uma mudança, por parte das plataformas, de uma abordagem de “acesso por solicitação” para “divulgação por padrão”, com as necessárias salvaguardas de privacidade, permitiria aos pesquisadores avaliar adequadamente os danos sociais gerados pela poluição da informação.

EMPODERAMENTO DO USUÁRIO

Grupos da sociedade civil e acadêmicos realizaram extensas pesquisas sobre a melhor forma de combater a informação falsa, a desinformação e o discurso de ódio, protegendo a liberdade de expressão. Vários enfatizaram a necessidade de soluções que envolvam os usuários e suas comunidades, e que ofereçam treinamento aos usuários da Internet sobre como limitar o impacto dos danos em suas próprias comunidades, descentralizando, assim, o poder das mãos das plataformas. Os usuários das plataformas, incluindo grupos marginalizados, devem ser encorajados, incluídos e envolvidos nesse espaço político. Os jovens, em particular, têm uma riqueza e profundidade de conhecimento. Como nativos digitais, os jovens, em particular mulheres e crianças, já são frequentemente alvos

de informação falsa e do discurso de ódio e, serão diretamente afetados por plataformas novas e emergentes. Os usuários mais jovens podem falar por experiência sobre o impacto diferenciado de várias propostas e suas possíveis falhas. Eles também contribuíram ativamente para a defesa on-line e os esforços de verificação de fatos.⁵¹

Habilidades aprimoradas de pensamento crítico podem tornar os usuários mais resilientes contra a manipulação digital. Especificamente, a alfabetização digital ensina os usuários a avaliar melhor as informações que encontram on-line e a transmiti-las de forma responsável. Uma série de entidades das Nações Unidas tiveram experiências valiosas neste campo. A campanha “Verificado” das Nações Unidas⁵² implementou com sucesso uma variedade de táticas, incluindo o foco em mensagens para usuário e pré-bunking – alertando os usuários sobre mentiras antes que eles se deparassem com elas, e campanhas de alfabetização digital

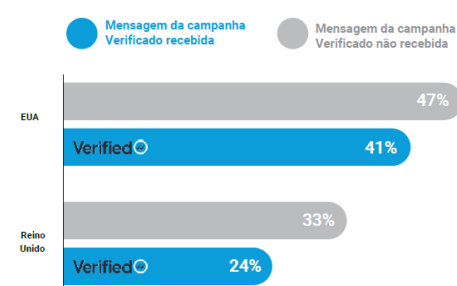
DESISTÍMULOS

Os atuais modelos de negócios da maioria das plataformas digitais priorizam o engajamento acima dos direitos humanos, da privacidade e da segurança. Isso inclui a monetização de dados pessoais para fins lucrativos, apesar das crescentes evidências de danos sociais causados por esse modelo de negócio.

FIGURA III

CAMPANHAS EFETIVAS DAS NAÇÕES UNIDAS DE COMBATE ÀS INFORMAÇÕES FALSAS E À DESINFORMAÇÃO

Probabilidade de compartilhamento de notícias falsas (2021)^a



^a Baseado na pesquisa conduzida pelo Instituto de Tecnologia de Massachusetts em março de 2021.

Alguns grupos da sociedade civil e pesquisadores exploraram caminhos para desmonetizar e, portanto, desincentivar a criação e disseminação de informações falsas, da desinformação e do discurso de ódio on-line, observando que, embora a liberdade de expressão seja um direito humano fundamental, lucrar com isso não é.⁵³ As propostas buscam abordar a lucratividade da desinformação, garantir total transparência em relação à monetização de conteúdo e das avaliações independentes de risco, além de desincentivar os envolvidos em publicidade on-line de permitir a desinformação.

Marcas que anunciam junto com informação falsa e o discurso de ódio correm o risco de prejudicar a eficácia de suas campanhas e, em última análise, prejudicar suas reputações. Os anunciantes podem desenvolver políticas claras a fim de evitar o financiamento inadvertido e legitimar a falta de informação, a desinformação e o discurso de ódio, além de ajudar a torná-los não lucrativos. As medidas de implementação podem incluir o gerenciamento de listas atualizadas de inclusão e exclusão e o uso de ferramentas de verificação de anúncios. Os anunciantes também podem pressionar as plataformas digitais a intensificar as ações para proteger a integridade da informação e podem abster-se de anunciar em meios de comunicação que alimentam o ódio e espalham informações falsas.⁵⁴

MÍDIA INDEPENDENTE

Novas medidas em dezenas de países continuam prejudicando a liberdade de imprensa. De acordo com o relatório global da UNESCO para 2022 de sua série sobre Tendências Mundiais em Desenvolvimento da Liberdade de Expressão e de Imprensa, 85% da população mundial experimentou um declínio na liberdade de imprensa em seus países no decorrer dos últimos cinco anos.⁵⁵ Com 2,7 bilhões de pessoas ainda offline,⁵⁶ outra prioridade é fortalecer a mídia independente, aumentar a prevalência de

iniciativas de checagem de fatos e sustentar reportagens confiáveis e precisas de interesse público. O debate público real depende dos fatos, contados com clareza e relatados de forma ética e independente. Repórteres éticos, com formação e condições de trabalho de qualidade, têm competências para restabelecer o equilíbrio Diante da desinformação. Eles são capazes de oferecer um serviço vital: informações precisas, objetivas e confiáveis sobre questões que importam.

À PROVA DO FUTURO

Mesmo enquanto se buscam soluções para proteger a integridade da informação no cenário atual, devemos garantir que as recomendações sejam à prova dos desafios futuros, abordando as tecnologias emergentes e as que ainda estão por vir.

Lançada em novembro de 2022, a plataforma ChatGPT-3 da Open AI conquistou 100 milhões de usuários até janeiro de 2023, tornando-se o aplicativo de consumo que mais cresce na história,⁵⁷ com muitas outras empresas correndo para desenvolver ferramentas competitivas. Embora tenha um potencial quase inimaginável para enfrentar os desafios globais, há preocupações sérias e urgentes sobre o potencial igualmente poderoso dos avanços recentes em inteligência artificial – incluindo geradores de imagens e deepfakes de vídeo – que ameaçam a integridade da informação. Relatórios e pesquisas recentes mostraram que ferramentas geradas de inteligência artificial produziram informação falsa, desinformação e discurso de ódio, apresentados de forma convincente aos usuários como se fossem reais.⁵⁸

O enviado especial da ONU para Tecnologia está liderando esforços para avaliar as implicações da inteligência artificial generativa e outras plataformas emergentes. Ao fazer isso, devemos aprender com os erros do passado. As

plataformas digitais foram lançadas no mundo sem consciência ou avaliação suficiente dos danos potenciais para sociedades e indivíduos. Temos agora a oportunidade de garantir que a história não se repita com a tecnologia emergente. A era da filosofia de “mover-se super rápido e lançar tudo novo” do Vale do Silício deve chegar ao fim. É essencial que a privacidade, segurança, transparência e segurança do usuário sejam integradas em todas as novas tecnologias e produtos desde o início.

RESPOSTAS DAS NAÇÕES UNIDAS

Medidas também estão sendo tomadas, inclusive pelas Forças de Paz e pelos escritórios nacionais das Nações Unidas, para monitorar, analisar e responder à ameaça que a informação falsa e a desinformação representa para o cumprimento dos mandatos das Nações Unidas. A [Estratégia e Plano de Ação das Nações Unidas sobre o Discurso de Ódio](#) estabelece orientações estratégicas para a Organização abordar o discurso de ódio nos níveis nacional e global. Em fevereiro de 2023, a UNESCO organizou a conferência “[Internet pela Confiança](#)” para discutir um conjunto de rascunhos de diretrizes globais para a regulamentação de plataformas digitais, que devem ser finalizadas ainda este ano.⁵⁹

Juntas, essas iniciativas e abordagens ajudam a apontar o caminho a seguir para os princípios subjacentes de um Código de Conduta das Nações Unidas.

Em direção a um Código de Conduta das Nações Unidas

O Código de Conduta das Nações Unidas para Integridade da Informação em Plataformas Digitais, que apresentarei na Cúpula do Futuro, está sendo estruturado a partir dos seguintes princípios:

- Compromisso com a integridade da informação
- Respeito aos direitos humanos
- Apoio à mídia independente
- Aumento da transparência
- Empoderamento dos usuários
- Fortalecimento da pesquisa e do acesso a dados
- Respostas ampliadas
- Desestímulos mais fortes
- Maior confiança e segurança

Esses princípios foram compilados das ideias centrais discutidas no presente resumo e estão alinhados e interligados com meu documento de sobre um Pacto Digital Global. Os Estados-membros serão convidados a implementar o Código de Conduta a nível nacional. As consultas continuarão com as partes interessadas para refinar ainda mais o conteúdo do Código de Conduta, bem como para identificar metodologias concretas para operacionalizar seus princípios.

O Código de Conduta pode basear-se nas seguintes recomendações:

Compromisso com a integridade da informação

- (a) Todas as partes interessadas devem abster-se de usar, apoiar ou ampliar a informação falsa e o discurso de ódio para qualquer finalidade, inclusive para alcançar objetivos políticos,

militares ou outros objetivos estratégicos, incitar a violência, minar os processos democráticos ou atingir populações civis, grupos vulneráveis, comunidades ou indivíduos.

Respeito aos direitos humanos

(b) Os Estados-membros devem:

(i) Assegurar que as respostas à informação falsa e ao discurso de ódio sejam consistentes com a lei internacional, incluindo a lei internacional de direitos humanos, e que não sejam utilizadas de forma a bloquear qualquer expressão legítima de pontos de vista ou opinião, inclusive por meio de desligamentos gerais da Internet ou proibições de plataformas ou meios de comunicação.

(ii) Implementar medidas regulatórias para proteger os direitos fundamentais dos usuários de plataformas digitais, incluindo mecanismos de fiscalização, com total transparência quanto às exigências impostas às empresas de tecnologia.

(c) Todas as partes interessadas devem cumprir os Princípios Orientadores sobre Empresas e Direitos Humanos.

Apoio à mídia independente

(d) Os Estados-membros devem garantir um cenário de mídia livre, viável, independente, plural e com forte proteção para jornalistas e para a mídia independente. Além disso, devem apoiar o estabelecimento, financiamento e treinamento de organizações independentes

de verificação de fatos nos idiomas locais.

- (e) A mídia jornalística deve assegurar relatórios independentes precisos e éticos, sustentados em treinamento de qualidade e condições de trabalho adequadas, de acordo com as normas e padrões internacionais de trabalho e direitos humanos.

Aumento da transparência

- (f) As plataformas digitais devem:

- (i) Garantir transparência significativa em relação aos algoritmos, dados, moderação de conteúdo e publicidade.

- (ii) Publicar e divulgar políticas acessíveis sobre informação falsa, desinformação e discurso de ódio, e relatar a prevalência de informações falsas coordenadas em seus serviços e a eficácia de políticas para combater tais questões.

- (g) A mídia jornalística deve garantir uma transparência substancial das fontes de financiamento e políticas de publicidade, além de distinguir claramente o conteúdo editorial da publicidade paga, inclusive ao publicar em plataformas digitais.

Empoderamento de usuários

- (h) Os Estados-membros devem garantir o acesso público a informações governamentais corretas, transparentes e de fontes confiáveis, particularmente informações que atendam ao interesse público, incluindo todos os aspectos dos Objetivos de Desenvolvimento Sustentável.

- (i) As plataformas digitais devem garantir o empoderamento e a proteção transparente do usuário, dando às pessoas um maior poder de escolha sobre o conteúdo que veem e como seus dados são usados. Devem permitir que os usuários comprovem

a identidade e a autenticidade livres de compensações de privacidade ou monetárias e estabeleçam processos transparentes de reclamação e denúncia do usuário, apoiados por mecanismos de revisão de reclamações independentes, bem divulgados e acessíveis.

- (j) Todas as partes interessadas devem investir em iniciativas robustas de alfabetização digital com a finalidade de treinar usuários de todas as idades para entender melhor como as plataformas digitais funcionam, como seus dados pessoais podem ser usados, bem como para identificar e responder à informação falsa, à desinformação e ao discurso de ódio. Uma atenção especial deve ser dada para garantir que jovens, adolescentes e crianças tenham pleno conhecimento de seus direitos nos espaços on-line.

Fortalecimento da pesquisa e acesso a dados

- (k) Os Estados-membros devem investir e apoiar pesquisas independentes sobre a prevalência e o impacto da informação falsa e do discurso de ódio em todos os países e idiomas, particularmente em contextos de maior vulnerabilidade e em idiomas diferentes do inglês, permitindo que a sociedade civil e a academia operem livremente e em segurança.

- (l) As plataformas digitais devem:

- (i) Permitir que pesquisadores e acadêmicos tenham acesso a dados, respeitando a privacidade do usuário. Os pesquisadores devem poder coletar amostras e dados qualitativos sobre indivíduos e grupos tidos como alvo pela informação falsa, desinformação e discurso de ódio para entender melhor o escopo e a natureza dos danos, respeitando a proteção de dados e os direitos humanos.

- (ii) Assegurar a ampla participação da sociedade civil nos esforços para lidar com

a informação falsa, a desinformação e o discurso de ódio.

Respostas ampliadas

(m) Todas as partes interessadas devem:

(i) Alocar recursos para combater e relatar as origens, disseminação e impacto das informações falsas, da desinformação e do discurso de ódio, respeitando as normas e padrões de direitos humanos e investindo amais em recursos de verificação de fatos em todos os países e contextos.

(ii) Formar amplas coalizões sobre integridade da informação, reunindo diferentes competências e abordagens que ajudem a preencher a lacuna entre organizações locais e empresas de tecnologia que operam em escala global.

(iii) Promover treinamentos e desenvolvimento de capacidades que visem ao entendimento de como as informações falsas, a desinformação e o discurso de ódio se manifestam, e ao fortalecimento das estratégias de prevenção e mitigação.

Desestímulo mais fortes

(n) As plataformas digitais devem se afastar de modelos de negócios que priorizem o engajamento acima dos direitos humanos, da privacidade e da segurança.

(o) Os anunciantes e as plataformas digitais devem garantir que os anúncios não sejam colocados ao lado de informações falsas, da desinformação ou do discurso de ódio online, e que a publicidade contendo informação falsa não seja promovida.

(p) A mídia de notícias deve garantir que toda publicidade paga e conteúdo publicitário sejam claramente marcados como tal e livres de informações falsas e do discurso de ódio.

Maior confiança e segurança

(q) As plataformas digitais devem:

(i) Garantir a segurança e a privacidade desde a concepção de todos os produtos, inclusive por meio de recursos adequados de confiança interna e experiência em segurança, juntamente com a aplicação consistente de políticas em todos os países e idiomas.

(ii) Investir em sistemas de moderação de conteúdo feitos por inteligência humana e artificial em todos os idiomas usados nos países de operação e garantir que os mecanismos de relatórios de conteúdo sejam transparentes, com uma taxa de resposta acelerada, especialmente em contextos de conflito.

(r) Todas as partes interessadas devem tomar medidas urgentes e imediatas para garantir o uso seguro, protetivo, responsável, ético e compatível com os direitos humanos da inteligência artificial, além de abordar as implicações dos avanços recentes neste campo para a disseminação de informações falsas e do discurso de ódio.

Próximos Passos

- O Secretariado das Nações Unidas realizará amplas consultas com partes interessadas sobre desenvolvimento do Código de Conduta das Nações Unidas, incluindo mecanismos de acompanhamento e implementação. Isso pode incluir o estabelecimento de um observatório independente composto por especialistas reconhecidos para avaliar as medidas tomadas pelos atores que se comprometerem com o Código de Conduta e outros mecanismos de denúncia.
- Para a manutenção e embasamento do Código, o Secretariado das Nações Unidas pode realizar estudos aprofundados para melhorar a compreensão da integridade da informação no âmbito global, especialmente em partes do mundo pouco pesquisadas.
- O secretário-geral estabelecerá capacidade dedicada no Secretariado das Nações Unidas para aumentar a resposta às informações falsas, à desinformação e ao discurso de ódio on-line que afetam a execução do mandato das Nações Unidas e de suas prioridades substantivas. Com base no monitoramento e na análise de especialistas, essa capacidade desenvolverá estratégias de comunicação sob medida para antecipar e/ou lidar rapidamente com as ameaças antes que se transformem em danos on-line e off-line e apoiar o treinamento dos funcionários das Nações Unidas e dos Estados-membros.
- Ações para apoiar os esforços dos Estados-membros, plataformas digitais e outras partes interessadas em aderir e implementar o Código de Conduta, quando finalizado.

Conclusão

Fortalecer a integridade da informação nas plataformas digitais é uma prioridade urgente para a comunidade internacional. Da saúde e igualdade de gênero à paz, justiça, educação e ação climática, as medidas que limitam o impacto das informações falsas, da desinformação e do discurso de ódio impulsionarão os esforços para alcançar um futuro sustentável, sem deixar ninguém para trás.

Mesmo com ações em nível nacional, esses problemas só podem ser totalmente resolvidos por meio de uma cooperação global mais forte. As ideias centrais descritas neste Informe de Política demonstram que o caminho para uma integridade da informação mais forte precisa ser baseado nos direitos humanos, e ser multissetorial e multidimensional.

Essas ideias serão elaboradas em uma série de princípios a serem considerados para um Código de Conduta das Nações Unidas para Integridade da Informação em Plataformas Digitais, com o objetivo de fornecer um modelo que reforce a integridade da informação, e, ao mesmo tempo, defenda vigorosamente os direitos humanos. Estou ansioso para colaborar com os Estados-membros e outras partes interessadas para transformar esses princípios em compromissos tangíveis.

Notas

- 1 [A/HRC/42/50; A/77/287; A/HRC/51/53](#); Nações Unidas, “Statement by Alice Wairimu Nderitu, Special Adviser on the Prevention of Genocide, condemning the recent escalation of fighting in Ethiopia”, press release, 19 de outubro de 2022; Office of the United Nations High Commissioner for Human Rights (OHCHR), “Myanmar: Social media companies must stand up to junta’s online terror campaign say UN experts”, comunicado de imprensa, 13 de março de 2023; OHCHR, “Freedom of speech is not freedom to spread racial hatred on social media: UN experts”, declaração, 6 de janeiro de 2023; Relatoria Especial para a Promoção e Proteção do Direito à Liberdade de Opinião e Expressão, “#JournalistsToo: women journalists speak out”, 24 de novembro de 2021; e OHCHR, “Sri Lanka: Experts dismayed by regressive steps, call for renewed UN scrutiny and efforts to ensure accountability”, press release, 5 de fevereiro de 2021.
- 2 [A/HRC/47/25](#), parágrafo 15.
- 3 Kalina Bontcheva e Julie Posetti, ed., Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression – Broadband Commission Research Report em “Freedom of Expression and Addressing Disinformation on the Internet” (Genebra, União Internacional de Telecomunicação (ITU); Paris, UNESCO, 2020).
- 4 Ver Nações Unidas, “Countering disinformation”, disponível em: www.un.org/en/countering-disinformation, e [A/77/287](#).
- 5 Disponível em www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_EN.pdf.
- 6 As Nações Unidas estão atualmente realizando um estudo para analisar as interligações e relações entre falta de informação, desinformação e discurso de ódio, e onde esses fenômenos relacionados, mas distintos, convergem e divergem nos níveis conceitual e operacional.
- 7 A Comissão Europeia define plataformas on-line em “Shaping Europe’s digital future: online platforms”, 7 de junho de 2022. Disponível em <https://digital-strategy.ec.europa.eu/en/policies/online-platforms>.
- 8 Veja OHCHR, “Moderating online content: fighting harm or silencing dissent?”, 23 de julho 2021.
- 9 Veja Nações Unidas, “Countering disinformation”, e [A/77/287](#).
- 10 [A/HRC/47/25](#).
- 11 Stephanie Kirchgassner e outros, “Revealed: the hacking and disinformation team meddling in elections”, The Guardian, 14 de fevereiro 2023.
- 12 Alexandre Alaphilippe e outros, “Doppelgänger – media clones serving Russian propaganda”, EU DisinfoLab, 27 de setembro de 2022.
- 13 United Nations Economist Network, “New economics for sustainable development: attention economy”.
- 14 Twitter, “About Twitter Blue”; e Meta, “Testing Meta Verified to help creators establish their presence”, 17 de março de 2023.
- 15 [A/77/287](#).
- 16 Em fevereiro de 2023, 173 Estados Membros eram Estados Partes do Pacto Internacional sobre Direitos Civis e Políticos.
- 17 As limitações à liberdade de expressão devem atender às seguintes condições bem estabelecidas: a legalidade, ou seja, as restrições devem ser previstas em lei de maneira que distinga com suficiente precisão as expressões lícitas e ilícitas; necessidade e proporcionalidade, ou seja, a limitação comprovadamente impõe o menor ônus ao exercício do direito e realmente protege, ou é susceptível de proteger, o legítimo interesse do Estado em questão; e legitimidade, ou seja, para serem legais, as restrições devem proteger apenas os interesses enumerados no artigo 19 (3) do Pacto Internacional sobre Direitos Civis e Políticos. These include audiovisual communications in Côte d’Ivoire, Morocco and Tunisia, and monitoring of incitement to violence by the United Nations Multidimensional Integrated Stabilization Mission in the Central African Republic.
- 18 Alekski Knuutila, Lisa-Maria Neudert e Philip N. Howard, “Who is afraid of fake news? Modeling risk perceptions of misinformation in 142 countries”, Harvard Kennedy School (HKS) Misinformation Review, vol. 3, No. 3 (Abril 2022).
- 19 ITU, [Measuring the Information Society](#) (Genebra, 2013).
- 20 Fundo das Nações Unidas para a Infância (UNICEF), “Protecting children online”, 23 de junho de 2022. Disponível em www.unicef.org/protection/violence-against-children-online.
- 21 UNESCO, artigos de trabalho sobre governança digital e os desafios para a confiança e segurança. Disponível em www.unesco.org/en/internet-conference/working-papers.
- 22 Ver Julie Posetti e Kalina Bontcheva, “Disinfodemic: deciphering COVID-19 disinformation” policy brief 1, (Paris, UNESCO, 2020), e “Disinfodemic: dissecting responses to COVID-19 disinformation” policy brief 2, (Paris, UNESCO, 2020).
- 23 Centro de Combate ao Ódio Digital, [Pandemic Profiteers: The Business of Anti-Vaxx](#) (2021).
- 24 Michael A Gisondi e outros, “A deadly infodemic: social media and the power of COVID-19 misinformation”, Journal of Medical Internet Research, vol. 24, No. 2 (Fevereiro 2022).
- 26 [A/77/288](#).
- 27 [A/77/287](#), parágrafo 6.
- 28 Em 2018, uma missão internacional independente de apuração de fatos nomeada pelo Conselho de Direitos Humanos declarou o Facebook como “a principal plataforma para discurso de ódio em Mianmar” ([A/HRC/42/50](#), parágrafo 72).

- 29 Ver Notícias das Nações Unidas, [“Hate speech: a growing, international threat”](#), 28 de janeiro de 2023, e [“Digital technology, social media fuelling hate speech like never before, warns UN expert”](#), 20 de outubro de 2022
- 30 Ver John Cook, “Deconstructing climate science denial”, em [Research Handbook in Communicating Climate Change](#), David C. Holmes e Lucy M. Richardson, eds. (Cheltenham, United Kingdom, Edward Elgar, 2020). Cook relatou que Abraham et al. (2014) resumiu como artigos contendo alegações negacionistas, como alegações de resfriamento em medições de satélite ou estimativas de baixa sensibilidade climática, foram refutadas de forma robusta na literatura científica. Similarmente, Benestad et al. (2016) tentou replicar descobertas em artigos contrários e encontrou uma série de falhas, como métodos estatísticos inapropriados, falsas dicotomias e conclusões baseadas em física mal concebida.
- 31 Global Witness, [“The climate divide: how Facebook’s algorithm amplifies climate disinformation”](#), 28 de março de 2022.
- 32 Analisado pelo Departamento de Comunicação Global, usando dados do Talkwalker.
- 33 Jeffrey A. Hicke e outros, “North America”, em Intergovernmental Panel on Climate Change, [Climate Change 2022: Impacts, Adaptation and Vulnerability](#), Contribuição do Grupo de Trabalho II ao Sexto Relatório de Avaliação do Painel Intergovernamental sobre Mudanças Climáticas (Cambridge, United Kingdom, Cambridge University Press, 2022).
- 34 Mei Li, Gregory Trencher e Jusen Asuka, [“The clean energy claims of BP, Chevron, ExxonMobil and Shell: a mismatch between discourse, actions and investments”](#), PLOS ONE, edição 17, No. 2 (Fevereiro 2022).
- 35 Robert J. Brulle e Carter Werthman, [“The role of public relations firms in climate change politics”](#), Climatic Change, vol. 169, No. 1–2 (Novembro 2021). De acordo com o Índice Global de Desinformação, um órgão de vigilância sem fins lucrativos, os anunciantes da indústria de tecnologia forneceram US\$36,7 milhões para 98 sites com desinformação climática em inglês em 2021. Um relatório de novembro de 2022 do Center for Countering Digital Hate, um grupo de campanha, revelou que, em Somente o Google, quase metade dos US\$ 23,7 milhões gastos em anúncios de busca por empresas de petróleo e gás nos últimos dois anos direcionaram termos de busca para sustentabilidade ambiental. A pesquisa do InfluenceMap encontrou 25.147 anúncios enganosos de 25 organizações do setor de petróleo e gás nas plataformas do Facebook nos Estados Unidos da América em 2020, com um gasto total de US\$ 9.597.376. Até agora, a resposta tem sido incomensurável com a escala do problema.
- 36 Ver Resolução da Assembleia Geral [76/227](#); resolução do Conselho de Direitos Humanos [49/21](#); e Ação Externa da União Europeia, [“Tackling disinformation, foreign information manipulation and interference”](#), 27 October 2021.
- 37 Lucina Di Meco, [“Monetizing misogyny: gendered disinformation and the undermining of women’s rights and democracy globally”](#), #ShePersisted, Fevereiro 2023.
- 38 Ver Andrew Puddephatt, [“Social media and elections”](#), Cuadernos de Discusión de Comunicación e Información, No. 14 (Montevideo, UNESCO, 2019); e Julie Posetti e outros, [“The chilling: global trends in online violence against women journalists”](#), trabalho de discussão de pesquisa (UNESCO, 2021).
- 39 EU Disinfo Lab, [“The role of “media” in producing and spreading disinformation campaigns”](#), 13 de outubro de 2021.
- 40 Ver Notícias das Nações Unidas, [“Social media poses ‘existential threat’ to traditional, trustworthy news: UNESCO”](#), 10 de março de 2022; e Anya Schiffrin e outros, [“Finding the funds for journalism to thrive: policy options to support media viability”](#), World Trends em Freedom of Expression and Media Development (Paris, UNESCO, 2022).
- 41 Comissão Europeia, [“Shaping Europe’s digital future: the 2022 Code of Practice on Disinformation”](#), 4 de julho de 2022.
- 42 Ver as observações do Vice-presidente da Comissão Europeia para Valores e Transparência, Věra Jourová, em European Commission, [“Code of Practice on Disinformation: new Transparency Centre provides insights and data on online disinformation for the first time”](#), notícias diárias, 9 de fevereiro de 2023. Disponível em https://ec.europa.eu/commission/presscorner/detail/en/mex_23_723.
- 43 Disponível em <https://unglobalcompact.org/library/2>.
- 44 Centro de Combate ao Ódio Digital e Campanha de Direitos Humanos, [“Digital hate: social media’s role in amplifying dangerous lies about LGBTQ+ people”](#), 10 de agosto de 2022.
- 45 Ver Andrew Puddephatt, [“Letting the sun shine in: transparency and accountability in the digital age”](#), World Trends em Freedom of Expression and Media Development (Paris, UNESCO, 2021).
- 46 Whose Knowledge?, Oxford Internet Institute and The Centre for Internet and Society, [State of the Internet’s Languages Report \(2022\)](#).
- 47 [A/HRC/38/35](#).
- 48 Billy Perrigo, [“Inside Facebook’s African sweatshop”](#), Time, 17 de fevereiro de 2022.
- 49 [A/HRC/42/50](#).
- 50 Exemplos notáveis incluem Maria Ressa, How to Stand up to a Dictator (New York, HarperCollins, 2022); e Max Fischer, The Chaos Machine (New York, Little, Brown and Company, 2022).
- 51 Ver UNICEF, [“Young reporters fact-checking COVID-19 information”](#).
- 52 Ver <https://shareverified.com/>.
- 53 O Índice Global de Desinformação, um grupo sem fins lucrativos, rastreia a publicidade colocada junto com a desinformação. As Nações Unidas foram vítimas dessa prática, com o Índice Global de Desinformação tendo encontrado anúncios da UNICEF colocados ao lado de artigos anti-vacinas e anúncios do Escritório do Alto Comissariado das Nações Unidas para Refugiados junto com conteúdo anti-refugiados.

- 54 Rede de Publicidade Consciente, manifestos. Disponível em www.consciousadnetwork.com/the-manifestos/.
- 55 UNESCO, *Journalism is a Public Good: World Trends in Freedom of Expression and Media Development – Global Report 2021/2022* (Paris, 2022).
- 56 ITU, “[Facts and figures 2021: 2.9 billion people still offline](#)”, 29 de novembro de 2021. The [global digital compact](#) to be taken up by Member States at the Summit of the Future, to be held in 2024, will outline shared principles for an open, free and secure digital future for all (ver www.un.org/techenvoy/global-digital-compact).
- 57 Krystal Hu, “[ChatGPT sets record for fastest-growing user base – analyst note](#)”, Reuters, 2 de fevereiro de 2023.
- 58 Ver Centro de combate ao ódio digital, “[Misinformation on Bard, Google’s new AI chat](#)”, 5 de abril de 2023; e Tiffany Hsu e Stuart A. Thompson, “[Disinformation researchers raise alarms about A.I. chatbots](#)”, The New York Times, 13 de fevereiro de 2023.
- 59 As diretrizes do projeto estão disponíveis em www.unesco.org/en/internet-conference

