



ADVOCACIA-GERAL DA UNIÃO
SECRETARIA-GERAL DE CONTENCIOSO

Excelentíssimo Senhor Ministro DIAS TOFFOLI, Relator do Recurso Extraordinário nº 1.037.396
(Repercussão Geral - **Tema 987**)

Repercussão Geral. Julgamento conjunto dos Temas 533 e 987.

Proposta de tese pelo Min. Relator, no Tema 987, consistente em Decálogo contra a violência digital e a desinformação, com o estabelecimento de **deveres anexos** de segurança, prevenção, cuidado e transparência, dirigidos aos provedores de aplicações de internet.

Uso indevido da **Inteligência Artificial**. Risco para a credibilidade de relevantes ações e programas governamentais. Ofensa à honra e à imagem de figuras públicas. Impacto negativo para a reputação e para a confiança dos administrados em relação à Administração Pública.

Requerimento a essa Egrégia Suprema Corte de que, ao formular a tese de Repercussão Geral, considere que os deveres anexos direcionados aos provedores de aplicações de internet também se aplicam aos casos de uso de inteligência artificial.

A **UNIÃO**, representada por seu Advogado-Geral^[1], nos autos do recurso extraordinário em epígrafe, interposto por FACEBOOK SERVIÇOS ONLINE DO BRASIL LTDA, vem, na qualidade de *amicus curiae*, informar e requerer conforme segue.

1. DO CONTEXTO ATUAL DOS AUTOS

1. Trata-se do exame conjunto de recursos extraordinários, com repercussão geral reconhecida, no qual se discute a interpretação de dispositivos do Marco Civil da Internet - Lei Federal nº 12.965, de 2014 -, notadamente quanto à extensão do **regime de responsabilidade dos provedores de aplicação por conteúdo ilícito gerado por terceiros**.

2. As questões em exame foram assim delimitadas por essa Suprema Corte, respectivamente: a) "*Discussão sobre a constitucionalidade do art. 19 da Lei nº 12.965/2014 que*

determina a necessidade de prévia e específica ordem judicial de exclusão de conteúdo para a responsabilização civil de provedor de internet, websites e gestores de aplicativos de redes sociais por danos decorrentes de atos ilícitos praticados por terceiros" (Tema 987) e; b) "dever de empresa hospedeira de sítio na internet fiscalizar o conteúdo publicado e de retirá-lo do ar quando considerado ofensivo, sem intervenção do Judiciário" [a partir de notificação extrajudicial] (Tema 533).

3. A União foi admitida na condição de *amicus curiae*, por meio do despacho e-Doc. 309, considerando a singularidade dos interesses então defendidos, a atuação proativa relativamente à matéria posta em julgamento, bem como em razão da relevância do tema.

4. De fato, o ente central apresentou informações iniciais por meio da petição e-Doc. 285, posteriormente complementadas através de memoriais (e-Doc. 330 e 344), além de ter requerido a juntada de material produzido em audiência pública realizada no âmbito da Advocacia-Geral da União, pertinente ao tema ora em análise (e-Doc. 354).

5. Por meio das mencionadas petições, a União manifestou-se no seguinte sentido:

a) a **declaração de inconstitucionalidade parcial sem redução do texto, para fins de se conservar a eficácia do art. 19 da Lei Federal n. 12.965/2014**, especificamente quanto à responsabilidade dos provedores de aplicação, tão somente nas hipóteses da liberdade de imprensa e do livre exercício da atividade jornalística, de qualquer espécie, no campo da licitude, **bem como da proteção da honra de agentes ou personalidades públicos**, além de disputas entre particulares relacionadas à reputação ou a direitos da personalidade".

b) Desnecessidade de ordem judicial específica ou de notificação extrajudicial para indisponibilização de conteúdo ilícito – exigência do dever de prevenção e precaução permanente e direta, sob pena de responsabilização objetiva, ante o risco da atividade - consideram-se conteúdos ilícitos, exemplificativamente, os seguintes":

(b.i) crimes contra o Estado Democrático de Direito, tipificados nos artigos 359-L a 359-R Código Penal, e a incitação ou apologia desses crimes, nos termos dos artigos 286 e 287 do Código Penal; atos de terrorismo e preparatórios de terrorismo, tipificados nos artigos 2º a 6º da Lei nº 13.260/2016; crimes praticados contra crianças e adolescentes, tipificados nos artigos 239 a 241-E da Lei nº 8.069/1990 (Estatuto da Criança e do Adolescente - ECA) e demais crimes previstos no Código Penal, e a incitação ou apologia desses crimes; crimes contra a saúde pública tipificados nos artigos 276, 277, 278 e 283 do Código Penal; crime de discriminação ou preconceito em razão de raça, cor, etnia, religião ou procedência nacional, tipificado no art. 20 da Lei nº 7.716/1989; e **a divulgação de fatos notoriamente inverídicos ou gravemente descontextualizados que atinjam a integridade do processo eleitoral**; crime de induzimento, instigação ou auxílio a suicídio ou a automutilação, tipificado no art. 122 do CP;

(b.ii) **conteúdos que violem: a proteção de crianças e adolescentes**, conforme disposto na Lei nº 8.069/1990 (Estatuto da Criança e do Adolescente - ECA); **as normas de proteção e a defesa do consumidor**, nos termos da Lei nº 8.078/1990 (Código de Defesa do Consumidor), especialmente, **os direitos de proteção contra publicidade enganosa e abusiva**, bem como da vida, saúde e segurança, e, por fim, **a veiculação de conteúdo ilícito publicitário, impulsionado, amplificado, recomendado, remunerado ou monetizado** - sob a concessão de qualquer vantagem econômica como retribuição à pessoa titular de conta, canal ou perfil que crie ou distribua conteúdo – a configurar casos de intermediação ativados provedor de aplicações de internet, mediante ato próprio; incluindo a disciplina normativa de anúncios e vendas de produtos e serviços de mercados regulados, a exemplo dos regulamentos de certificação e registros de produtos da ANVISA, ANATEL e IBAMA;

c) Desnecessidade de ordem judicial específica para adoção de medidas em face de conteúdo ilícito, com possibilidade de responsabilização apenas em caso de omissão após o recebimento de notificação extrajudicial da vítima ou interessado: **fraudes na invasão de perfis e contas ou criação de perfis e contas fraudulentos em nome de terceiros**, ressalvada a possibilidade de criação de perfis e contas para fins lícitos e compatíveis com a liberdade de expressão, seja na forma de crítica, sátira ou paródia; **manifesta e deliberada desinformação em matéria de políticas públicas e de legitimação de função pública**, sem prejuízo da

moderação de conteúdo por iniciativa própria das empresas, por violação a lei ou seus termos de uso, e do regime inerente às hipóteses previstas no item acima".

d) Tendo em vista a proteção da liberdade de expressão, a Advocacia-Geral da União reforça a necessidade de cumprimento dos deveres decorrentes da legislação brasileira, entre os quais o de garantir a informação adequada e clara, a transparência, a tutela da expectativa de continuidade do serviço, além de assegurar o devido processo no procedimento de moderação, para observar, ao menos:

i) notificação da medida de moderação aplicada - como remoção ou rotulagem de conteúdo, conforme o caso, indisponibilização de canal, conta ou perfil - ainda que diferida em casos graves;

ii) fundamentação que aponte as razões para aplicação e o conteúdo ou a conta que deu causa à decisão, de maneira clara e específica, em língua portuguesa;

iii) forma e prazos para exercício do direito de pedir revisão;

iv) se a decisão foi tomada exclusivamente por meio de sistemas automatizados, fornecimento de informações claras e adequadas a respeito dos critérios utilizados para a decisão, nos termos do art. 20, § 1º, da Lei nº 13.709/2018, quando cumpridos os requisitos para tanto;

v) resposta de modo fundamentado e objetivo aos pedidos de revisão, em prazo razoável, adotando sua reversão imediata em caso de equívoco;

vi) preservação de elementos de prova, em caso de violação a direitos, acompanhados dos respectivos metadados associados ao conteúdo ou perfil tornado indisponível;

vii) divulgação de relatórios de transparência periódicos sobre os procedimentos de decisão, tanto pela moderação como pela manutenção de conteúdos, acompanhados do tempo de resposta, com informações quantitativas e agregadas por operação, em formato aberto; e

viii) disponibilização de repositórios de publicidade digital permanentemente atualizados, contendo a íntegra dos anúncios veiculados, acompanhada de informações precisas sobre a identidade do anunciante, os valores pagos pela veiculação e as características gerais da audiência contratada.

6. O julgamento de mérito teve início em 27.11.2024, ocasião em que o Min. DIAS TOFFOLI (relator do **Tema 987**), proferiu voto pelo desprovimento do recurso extraordinário e a decretação da inconstitucionalidade do art. 19 do Marco Civil da Internet, fixando-se como regra a responsabilização dos provedores de aplicação de internet, nos termos do art. 21 do mesmo diploma. Em acréscimo, concluiu que os referidos provedores respondem civilmente de forma objetiva e independentemente de notificação, pelos danos decorrentes de conteúdos gerados por terceiros, além de possuírem uma série de **deveres anexos** de segurança, prevenção, cuidado e transparência.

7. Ao final do seu voto, o Min. Relator propôs tese de Repercussão Geral consistente em **Decálogo contra a violência digital e a desinformação**. Eis o teor da tese então proposta:

1. É inconstitucional o art. 19, caput e § 1º, do Marco Civil da Internet, sendo inconstitucionais, por arrastamento, os demais parágrafos do art. 19;

2. Como regra geral, o provedor de aplicações de internet será responsabilizado civilmente, nos termos do art. 21 do MCI, pelos danos decorrentes de conteúdos gerados por terceiros, inclusive na hipótese de danos à intimidade, à vida privada, à honra e à imagem, quando, notificado pelo ofendido ou seu representante legal, preferencialmente pelos canais de atendimento, deixar de promover, em prazo razoável, as providências cabíveis, ressalvadas as disposições da legislação eleitoral e os atos normativos expedidos pelo TSE;

2.1. Em todos os casos, a apuração da responsabilidade deverá levar em conta a(s) atividade(s) efetivamente desempenhada(s) pelo provedor de aplicações (art. 3º, inciso VI, do MCI) e o grau de interferência dessa(s) atividade(s), inclusive por atuação algorítmica e/ou automatizada, no fluxo informacional;

2.2. As plataformas e os blogs jornalísticos respondem exclusivamente na forma da lei específica (Lei nº 13.188, de 11 de novembro de 2015), a qual já foi declarada constitucional pelo Supremo Tribunal Federal (ADI nº 5.418, Rel. Min. Dias Toffoli, Tribunal Pleno, julgado em 11/3/21, DJe de 24/5/21).

3. O provedor de aplicações de internet, responde civilmente de forma objetiva e independentemente de notificação, pelos danos decorrentes de conteúdos gerados por terceiros, nas seguintes hipóteses:

3.1. quando recomendem, impulsionem (de forma remunerada ou não) ou moderem tais conteúdos, havendo responsabilidade solidária com o respectivo anunciante ou patrocinador, quando se tratar de anúncios publicitários ou de material patrocinado;

3.2. quando se tratar de conta inautêntica (também denominada de “perfil falso”), ou de conta desidentificada e/ou automatizada;

3.3. quando se tratar de direitos do autor e conexos, solidariamente com o terceiro responsável pela efetiva publicação/postagem do conteúdo, na forma dos arts. 102 a 104 da Lei nº 9.610, de 1998;

3.4. quando configurarem práticas previstas no seguinte rol taxativo:

(a) crimes contra o Estado Democrático de Direito (CP, art. 296, parágrafo único; art. 359-L, art. 359-M, art. 359-N, art. 359-P, art. 359-R);

(b) atos de terrorismo ou preparatórios de terrorismo, tipificados pela Lei nº 13.260, de 2016;

(c) crime de induzimento, instigação ou auxílio a suicídio ou à automutilação (CP, art. 122);

(d) crime de racismo (Lei nº 7.716, de 1989, arts. 20, 20-A, 20-B e 20-C);

(e) qualquer espécie de violência contra a criança, o adolescente e as pessoas vulneráveis, incluídos os crimes previstos nos arts. 217-A a 218-C do Código Penal, com redação dada pelas Leis nº 12.015, de 2009, e nº 13.718, de 2018, e na Lei nº 8.069, de 1990, e observada a Lei nº 13.257, de 2016, e a Res. CONANDA nº 245, de 2024;

(f) qualquer espécie de violência contra a mulher, incluindo os crimes da Lei nº 14.192, de 2021;

(g) infração sanitária, por deixar de executar, dificultar ou opor-se à execução de medidas sanitárias em situação de Emergência em Saúde Pública de Importância Nacional, nos termos do art. 10 da Lei nº 6.437, de 1977;

(h) tráfico de pessoas (CP, art. 149-A);

(i) incitação ou ameaça da prática de atos de violência física ou sexual (CP, art. 29 c/c arts. 121, 129, 213, 215, 215-A, 216-A, 250 e 251 c/c art. 147);

(j) divulgação de fatos notoriamente inverídicos ou gravemente descontextualizados que levem à incitação à violência física, à ameaça contra a vida ou a atos de violência contra grupos ou membros de grupos socialmente vulneráveis;

(k) divulgação de fatos notoriamente inverídicos ou descontextualizados com potencial para causar danos ao equilíbrio do pleito ou à integridade do processo eleitoral (Res. nº 23.610/2019, arts. 9-C e 9-D)

3.5. Havendo dúvida razoável acerca da configuração de uma das condutas mencionadas no item 3.4, aplica-se o regime do art. 21, na forma do item 2 desta tese;

4. O disposto nos itens 2 e 3 não se aplica:

(a) aos provedores de serviços de e-mail;

(b) aos provedores de aplicações cuja finalidade primordial seja a realização de reuniões fechadas por vídeo ou voz;

(c) aos provedores de serviços de mensageria instantânea (também chamadas de provedores de serviços de mensageria privada), exclusivamente no que concerne às comunicações interpessoais entre interlocutores certos e determinados, resguardadas pelo sigilo das comunicações (art. 5º, inciso XII, da CF/88);

5. Os provedores que funcionarem como marketplaces respondem objetiva e solidariamente com o respectivo anunciante nas hipóteses de anúncios de produtos de venda proibida ou sem certificação ou homologação pelos órgãos competentes no país (quando exigida), sem prejuízo da responsabilidade por vício ou defeito do produto ou serviço, conforme o Código de Defesa do Consumidor, e da aplicação do regime do art. 21 do MCI, na forma do item 2 desta tese, nas hipóteses residuais.

6. Os provedores de aplicações de internet devem atuar de forma responsável, transparente e cautelosa, a fim de assegurar um ambiente digital seguro, previsível e confiável, baseado nos princípios gerais da boa-fé, da função social e da prevenção e mitigação dos danos;

7. Os provedores de aplicações de internet possuem os seguintes **deveres anexos**:

(a) atualizar e manter atualizados os “termos e condições de uso” (ou documento equivalente) e os regulamentos adicionais, dando-lhes ampla publicidade;

(b) criar mecanismos para assegurar a autenticidade das contas e a correta identificação dos respectivos usuários, adotando as medidas necessárias para **obstar a criação de contas inautênticas, de contas desidentificadas ou automatizadas**, e para bloqueá-las, acaso sejam criadas;

(c) elaborar os respectivos códigos de conduta;

(d) estabelecer regras claras e procedimentos padronizados para a moderação de conteúdos, aos quais se dará ampla publicidade;

(e) proceder à constante atualização dos critérios e métodos empregados para a moderação de conteúdos, dando ampla publicidade aos usuários;

(f) **combater a difusão de desinformação e de notícias fraudulentas nos ambientes virtuais**, adotando as providências necessárias para a neutralização de redes artificiais de distribuição de conteúdo destinados a promovê-las, assim como para a identificação do perfil/conta do qual se originou a desinformação ou notícia fraudulenta para fins de encaminhamento dos dados às autoridades competentes para as providências cabíveis;

(g) monitorar os riscos sistêmicos de seus ambientes digitais, produzindo relatórios semestrais de transparência, dos quais constem os riscos identificados e as medidas preventivas e/ou de neutralização adotadas, aos quais se dará ampla publicidade;

(h) disponibilizar canais específicos de notificação, preferencialmente eletrônicos, para o recebimento de denúncias quanto à existência de conteúdo considerado infringente (com teor ofensivo ou ilícito), cuja apuração será prioritária;

(i) produzir relatórios semestrais de transparência relativamente à gestão e à resolatividade das reclamações pelos seus sistemas internos.

8. **Também é dever anexo** dos provedores de aplicações de internet o de atuar diligentemente para **prevenir e mitigar práticas ilícitas no âmbito de seus serviços, adotando as providências necessárias para combater a disseminação de conteúdos ilegais** que configurem as condutas indicadas taxativamente no subtópico 3.4 desta tese e para a identificação do perfil/conta do qual se originou a desinformação ou notícia fraudulenta, encaminhando esses dados às autoridades competentes para as providências cabíveis

9. Os provedores de aplicações de internet devem disponibilizar a usuários (e a não usuários) canais específicos de atendimento, preferencialmente eletrônicos, que sejam facilmente acessíveis, amplamente divulgados nas respectivas plataformas e permanentemente disponíveis, a fim de se possibilitar uma comunicação rápida e eficiente, bem como assegurar acesso temporário dos reclamantes a sistema interno pelo qual seja possível acompanhar o andamento da reclamação feita, as deliberações a respeito e as providências adotadas.

10. Os provedores de aplicações de internet com sede no exterior e atuação no Brasil devem constituir representante no país, cuja identificação e informações para contato deverão ser disponibilizadas e estar facilmente acessíveis nos respectivos sítios. Essa representação deve conferir ao representante, necessariamente pessoa jurídica com sede no país, plenos poderes para (a) responder perante as esferas administrativa e judicial; (b) prestar às autoridades competentes informações relativas ao funcionamento do provedor, às regras e aos procedimentos utilizados para moderação de conteúdo e para gestão das reclamações pelos sistemas internos; aos relatórios de transparência, monitoramento e gestão dos riscos sistêmicos; às regras para o perfilamento de usuários (quando for o caso) e para a veiculação de publicidade; (c) cumprir as determinações judiciais; e (d) responder e cumprir eventuais penalizações, multas e afetações financeiras em que o representado incorrer, especialmente por descumprimento de obrigações legais e judiciais.

8. Ato contínuo, na sessão de 11.12.2024, o Min. LUIZ FUX (relator do **Tema 533**) proferiu seu voto para negar provimento ao recurso extraordinário, mas propondo a manutenção do dispositivo impugnado, mediante tese no sentido de que:

1. A disposição do art. 19 do Marco Civil da Internet (Lei Federal nº 12.965/2014) não exclui a **possibilidade de responsabilização civil de provedores de aplicações de internet por conteúdos gerados por terceiros** nos casos em que, tendo ciência inequívoca do cometimento de atos ilícitos, seja porquanto evidente, seja porque devidamente informados por qualquer meio idôneo, não procederem à remoção imediata do conteúdo.

2. Considera-se evidentemente ilícito (item 1) o conteúdo gerado por terceiro que veicule discurso de ódio, racismo, pedofilia, incitação à violência, apologia à abolição violenta do Estado Democrático de Direito e apologia ao Golpe de Estado. Nestas hipóteses específicas, há para as empresas provedoras um dever de monitoramento ativo, com vistas à preservação eficiente do Estado Democrático de Direito.

3. Nos casos de postagens ofensivas à honra, à imagem e à privacidade de particulares, a ciência inequívoca da ilicitude por parte das empresas provedoras, necessária à responsabilização civil, dependerá de sua prévia e fundamentada notificação pelos interessados, que poderá ser realizada por qualquer meio idôneo, cabendo às plataformas digitais o dever de disponibilizar meios eletrônicos eficientes, funcionais e sigilosos para o recebimento de denúncias e reclamações de seus usuários que se sintam lesados.

4. **É presumido, de modo absoluto, o efetivo conhecimento da ilicitude do conteúdo produzido por terceiros por parte da empresa provedora de aplicações de internet, nos casos de postagens onerosamente impulsionadas**”

9. Na sessão de 18.12.2024, o Min. LUÍS ROBERTO BARROSO, por sua vez, antecipou seu voto, anuindo com a premissa de que uma compreensão literal do art. 19 do Marco Civil da Internet *"não dá proteção suficiente a direitos fundamentais, como a dignidade da pessoa humana, e a valores importantes para a democracia"*. No entanto, apresentou solução hermenêutica com nuances diversas, no sentido de que as plataformas *"devem ser responsabilizadas por conteúdos de terceiros caso deixem de tomar as providências necessárias para remover postagens com teor criminoso"*.

10. A seu ver, em complemento, no lugar de um "*monitoramento ativo, com responsabilidade, independentemente de notificação, por cada conteúdo individual, as empresas devem ter o chamado dever de cuidado, ou seja, **devem trabalhar para reduzir os riscos sistêmicos criados ou potencializados por suas plataformas***". As medidas, a seu ver, devem minimizar esses riscos e seus impactos negativos sobre direitos individuais e coletivos, segurança e estabilidade democrática.

11. Posteriormente, o julgamento foi suspenso em virtude de pedido de vista formulado pelo Min. ANDRÉ MENDONÇA.

12. No que aqui importa, extrai-se do voto do Min. Relator DIAS TOFFOLI ter sido evidenciada a **necessidade de se conferir maior atenção quanto ao uso da inteligência artificial** nas redes sociais, notadamente nos casos de recomendação, impulsionamento ou moderação de conteúdos:

Assim, em se tratando de dano decorrente de atuação humana, por *sponte* própria ou atendendo a alguma reclamação específica, **de atuação algorítmica ou automatizada (incluída a eventualmente proveniente do uso de ferramentas de inteligência artificial) para a recomendação, o impulsionamento (remunerado ou não) ou a moderação de conteúdos, da qual resulte a remoção de conteúdo ou a suspensão ou bloqueio de usuários, os provedores de aplicações poderão responder em conformidade com as regras existentes no ordenamento jurídico brasileiro como um todo.** (grifou-se)

13. Demais disso, o Min. Relator também revelou preocupação quanto à utilização de *deep fakes*, concluindo por seus efeitos deletérios para a sociedade, sobretudo no âmbito da política:

O que dizer, então, da desinformação que possui formato de imagem, de áudio ou vídeo? Da **desinformação produzida por meio de ferramentas de inteligência artificial** capazes de criar registros (quase) perfeitos de pessoas reais fazendo ou dizendo algo que nunca fizeram ou disseram? Isso já existe e tem nome: **deep fake**. Não precisamos fazer muito esforço para concluir que **seus efeitos podem ser desastrosos, principalmente no campo da política**. Afinal, se decisões políticas são mais emocionais que racionais, e se “uma imagem vale por mil palavras”, como reverter a convicção (política) formada em cidadãos comuns a partir de deep fakes?

14. Ainda sobre a **inteligência artificial**, o Min. Relator enfatizou que, "*Ao adotar essa postura ativa, promovendo a recomendação e o impulsionamento (remunerado ou não) dos conteúdos que circulam por seus ecossistemas, com emprego de algoritmos supercomplexos e ferramentas de inteligência artificial, ou, ainda, ao moderar conteúdos, seja por atuação humana seja de modo automatizado, *sponte* própria ou atendendo a reclamação específica, esses provedores interferem de forma preponderante no fluxo informacional, devendo responder pelos respectivos atos*".

15. Ao final do voto, o Min. DIAS TOFFOLI propôs tese de Repercussão Geral consistente em "Decálogo contra a violência digital e a desinformação", no qual foram especificados **deveres anexos** a serem seguidos pelos provedores de aplicações de internet, sendo oportuno aqui novamente destacá-los:

7. Os provedores de aplicações de internet possuem os seguintes **deveres anexos**:

- (a) atualizar e manter atualizados os "termos e condições de uso" (ou documento equivalente) e os regulamentos adicionais, dando-lhes ampla publicidade;
- (b) **criar mecanismos para assegurar a autenticidade das contas e a correta identificação dos respectivos usuários, adotando as medidas necessárias para obstar a criação de contas inautênticas, de contas desidentificadas ou automatizadas**, e para bloqueá-las, acaso sejam criadas;
- (c) elaborar os respectivos códigos de conduta;
- (d) estabelecer regras claras e procedimentos padronizados para a moderação de conteúdos, aos quais se dará ampla publicidade;
- (e) proceder à constante atualização dos critérios e métodos empregados para a moderação de conteúdos, dando ampla publicidade aos usuários;
- (f) **combater a difusão de desinformação e de notícias fraudulentas nos ambientes virtuais, adotando as providências necessárias para a neutralização de redes artificiais de distribuição de conteúdo destinados a promovê-las, assim como para a identificação do perfil/conta do qual se originou a desinformação ou notícia fraudulenta para fins de encaminhamento dos dados às autoridades competentes para as providências cabíveis**;
- (g) monitorar os riscos sistêmicos de seus ambientes digitais, produzindo relatórios semestrais de transparência, dos quais constem os riscos identificados e as medidas preventivas e/ou de neutralização adotadas, aos quais se dará ampla publicidade;
- (h) disponibilizar canais específicos de notificação, preferencialmente eletrônicos, para o recebimento de denúncias quanto à existência de conteúdo considerado infringente (com teor ofensivo ou ilícito), cuja apuração será prioritária;
- (i) produzir relatórios semestrais de transparência relativamente à gestão e à resolutividade das reclamações pelos seus sistemas internos.

8. **Também é dever anexo** dos provedores de aplicações de internet o de atuar diligentemente para **prevenir e mitigar práticas ilícitas no âmbito de seus serviços, adotando as providências necessárias para combater a disseminação de conteúdos ilegais** que configurem as condutas indicadas taxativamente no subtópico 3.4 desta tese e para a identificação do perfil/conta do qual se originou a desinformação ou notícia fraudulenta, encaminhando esses dados às autoridades competentes para as providências cabíveis

16. Nas palavras do Min. Relator, "*Também é dever anexo dos provedores de aplicações de internet o de atuar diligentemente para prevenir e mitigar práticas ilícitas no âmbito de seus serviços, adotando as providências necessárias para combater a disseminação de conteúdos ilegais gerados por terceiros que configurem as práticas indicadas taxativamente no subtópico 7.2.3 e para a identificação do perfil/conta do qual se originou a desinformação ou notícia fraudulenta, encaminhando esses dados às autoridades competentes para as providências cabíveis*".

17. **Contudo, a despeito da relevância do tema da inteligência artificial para o presente julgamento e da rápida evolução da tecnologia - contendo implicações diretas para a responsabilização dos provedores de aplicações de internet -, observa-se que tal abordagem não constou expressamente na proposta de tese de Repercussão Geral do Ministro relator.**

2. USO INDEVIDO DA INTELIGÊNCIA ARTIFICIAL. RISCO PARA AÇÕES GOVERNAMENTAIS. OFENSA À HONRA E À IMAGEM DE FIGURAS PÚBLICAS. NECESSIDADE DE PREVISÃO NA TESE DE REPERCUSSÃO GERAL NO SENTIDO DE QUE OS DEVERES ANEXOS TAMBÉM SE APLICAM AOS CASOS DE USO DE I.A.

18. A preocupação com a temática da inteligência artificial, do uso indevido da imagem de figuras públicas e do consequente prejuízo para políticas públicas governamentais já estava contida na manifestações anteriormente apresentadas pela União, ao defender a responsabilização das plataformas digitais pelo **conteúdo ilícito** ali publicado.

19. Na petição e-Doc. 285, a Advocacia-Geral da União aduziu que, *"ao mesmo tempo em que este **novo ambiente digital** facilita a livre manifestação de pensamento e o compartilhamento de informações, ele coloca em risco outros direitos fundamentais, como a **proteção à honra**, à privacidade e à **imagem**, estabelecidos no art. 5º, inciso X, da Constituição Federal"*.

20. Por sua vez, no memorial e-Doc. 330, a União defendeu a *"**desnecessidade de ordem judicial específica para adoção de medidas em face de conteúdo ilícito**, com possibilidade de responsabilização apenas em caso de omissão após o recebimento de notificação extrajudicial da vítima ou interessado"*, em especial para os casos de *"**manifesta e deliberada desinformação em matéria de políticas públicas e de legitimação de função pública**"*.

21. No mesmo sentido, no memorial complementar e-Doc. 344, a União argumentou pela conservação da eficácia do art. 19 da Lei Federal n. 12.965/2014, **em especial para a "proteção da honra de agentes ou personalidades públicas"**.

22. Já na petição e-Doc. 354, a União defendeu a necessidade de serem definidas *"**balizas seguras para a responsabilização dos provedores de aplicações por danos decorrentes de atos ilícitos praticados por terceiros**"*, para que *"**se promova um ambiente digital seguro e caracterizado pelo respeito aos direitos fundamentais e aos valores democráticos**"*.

23. Oportuno ressaltar, também, as seguintes passagem constantes na petição e-Doc. 354, na qual a União destacou trechos da fala de representante da empresa Meta durante a audiência pública então realizada:

Na audiência pública de 28.03.2023, por exemplo, **seu representante destacou os investimentos que vinham sendo feitos na direção contrária, isto é, a fim de se aprimorar o combate a conteúdos nocivos**, incitação à violência e discurso de ódio no país - restando claro que existem ferramentas tecnológicas para tanto, **inclusive mediante o emprego de inteligência artificial** - com relato de expressivos números quanto a conteúdos nocivos. Confira-se:

"(...) a Meta investe bilhões de dólares e trabalha no desenvolvimento de times de tecnologia, **inclusive, inteligência artificial**, para fazer valer os termos de uso e as políticas. A empresa tem um time com milhares de colaboradores dedicados à segurança e à integridade dos apps, incluindo pessoas no Brasil.

Falando sobre o primeiro turno das eleições, **a Meta removeu 135 mil anúncios de natureza eleitoral e mais de três milhões de conteúdos no Facebook e no Instagram por violação das políticas** que vedam conteúdo violento, incitação à violência e discursos de ódio. Esses três milhões são para o Brasil, no período entre agosto de 2022 e janeiro de 2023.

Evidentemente, reconhecemos que mais pode ser feito pelas plataformas. Diante da escala dos nossos serviços, proibir determinados conteúdos não vai significar incidência zero desses conteúdos, mas **a Meta está comprometida a seguir aprimorando esses sistemas**. Hoje, obtemos resultados muito superiores aos que obtínhamos na época dos fatos narrados nos recursos em análise"

24. Nesse cenário, observa-se que **o uso indevido da inteligência artificial para a produção de conteúdo ilícito ainda é um desafio a ser enfrentado no presente julgamento**, sendo fundamental traçar balizas seguras para a responsabilização dos provedores de aplicação no que tange ao uso da referida tecnologia.

25. Para ilustrar a relevância e a atualidade do tema, a União destaca o **estudo realizado pela entidade acadêmica NetLab** - Laboratório de Estudos de Internet e Redes Sociais da Universidade Federal do Rio de Janeiro (UFRJ) -, o qual publicou pesquisa intitulada "*Danos causados pela publicidade enganosa na Meta: Anúncios fraudulentos promovem desinformação sobre o Pix para lesar cidadãos brasileiros*^[2]". No referido estudo foram apresentadas evidências sobre como **anunciantes exploraram indevidamente políticas públicas** voltadas à inclusão financeira para aplicar golpes em cidadãos brasileiros por meio do impulsionamento de anúncios nas plataformas da Meta, **inclusive por meio da inteligência artificial**.

26. Na amostra objeto da pesquisa, especificamente entre 10 e 21 de janeiro de 2025, foram identificados **1.770 anúncios com conteúdo fraudulento** que promoviam golpes e informações falsas sobre valores a receber pela população e outros temas relacionados às novas regras de envio de informações de transações via Pix à Receita Federal.

27. As peças publicitárias fraudulentas analisadas ofertavam programas governamentais reais e fictícios, passando-se por páginas de instituições públicas ou privadas, **por meio da manipulação da imagem de lideranças políticas com Inteligência Artificial (IA)**, ao permitir a adulteração de falas ou a descontextualização de reportagens antigas para legitimar golpes.

28. De acordo com o aludido estudo, o alcance das fraudes foi **potencializado pela utilização das ferramentas de marketing** da empresa Meta, que permitem a compra de anúncios segmentados de acordo com critérios demográficos, geográficos, conforme interesses dos anunciantes.

29. Nesse sentido, os pesquisadores identificaram que *"70,3% dos anúncios analisados foram adulterados com Inteligência Artificial, ao passo que 40,5% foram feitos por páginas que se passavam pelo Governo Federal"*. Ainda, é pertinente ressaltar os seguintes dados apurados no estudo:

Ao todo, 718 anúncios (40,5% do total analisado), foram veiculados por páginas que se passavam por perfis oficiais do Governo Federal, seja por usarem nomes como "Governo Federal", "Governo do Brasil" e "GOV", seja por utilizarem em suas fotos de perfil logos e imagens associadas à administração federal ou a suas instituições.

(...)

Além destas páginas falsas, a credibilidade de instituições públicas federais também foi instrumentalizada por anúncios que são ilustrados por logos da Caixa Econômica Federal (9,1%, ou 162 anúncios), do Banco Central (17,8%, ou 315 anúncios) e da Receita Federal (15,5%, ou 275 anúncios).

30. De acordo com os pesquisadores, *"os anúncios eram frequentemente ilustrados por vídeos manipulados com uso de Inteligência Artificial de figuras públicas e autoridades brasileiras que 'instruíam' cidadãos a conferir e resgatar, via Pix, as quantias às quais eles supostamente teriam direito"*.

31. Com efeito, a falta de medidas de moderação, verificação, controle e transparência de anúncios propicia terreno fértil para a propagação de interesse criminoso em explorar a utilização da inteligência artificial, compondo a chamada **criminalidade cibernética**^[3], sobretudo em desfavor dos **extratos sociais mais vulneráveis**^[4], expondo os consumidores brasileiros a maiores riscos.

32. Diante da relevância dos dados obtidos pelo NetLab - com enfoque no descrédito gerado para diversas políticas públicas do Governo Federal e no uso indevido da imagem de figuras públicas^[5] -, a **Procuradoria Nacional da União de Defesa da Democracia da Procuradoria-Geral da União/AGU** ajuizou **Ação Civil Pública contra o "Facebook Brasil"**, no dia 29.04.2025, perante a Seção Judiciária do Distrito Federal^[6], a qual foi autuada sob o nº 1039882-92.2025.4.01.3400 e distribuída perante a 17ª Vara Federal Cível.

33. Na referida ação, a União requer tutela de urgência para que a ré adote *"medidas técnicas eficazes e específicas, compreendendo sistemas de bloqueio automático, a fim de impedir o*

uso indevido de símbolos e marcas de governo federal em anúncios e impulsionamentos", bem como utilize sistemas de bloqueio automático para "impedir a utilização indevida de imagens e vídeos, inclusive manipulados por inteligência artificial, de figuras públicas no conteúdo de anúncios e impulsionamento, notadamente autoridades federais como o Presidente da República, o Vice-Presidente da República e os Ministros de Estado", além da condenação por danos morais coletivos.

34. Observando a **realidade internacional**, vale destacar que nos Estados Unidos já é bilionário o montante auferido com as fraudes usando perfis falsos com supostas dicas de negócios, investimentos e soluções para endividamentos, inclusive se valendo das ferramentas de segmentação das plataformas para impactar as pessoas mais vulneráveis. O citado montante saltou de 42 milhões de dólares em 2017 para 1,2 bilhão de dólares em 2022.

35. No Brasil, por sua vez, mapeamento da OLX, em parceria com a AllowMe (<https://www.uol.com.br/tilt/noticias/redacao/2023/02/07/medo-de-comprar-online-veja-os-golpes-recentes-mais-aplicados.htm>), estima que brasileiros teriam perdido R\$ 551 milhões de reais em 2022 com fraudes online, as quais também utilizaram em grande parte o impulsionamento de anúncios fraudulentos via plataformas digitais, ou seja, com ganho financeiro para as empresas proprietárias dessas plataformas.

36. Nesse contexto, deve ser lembrado que **as plataformas digitais não são agentes passivos em relação aos conteúdos que nela transitam**, pois exercem atividade de mediação dos conteúdos exibidos para os seus usuários, definindo o que será exibido, o que pode ser moderado, o alcance das publicações, bem como o direcionamento das recomendações de conteúdos e de contas.

37. De fato, "ferramentas como chatbots, técnicas de deep fake e outras similares permitem a criação de textos, fotos, áudios e vídeos falsos com aparência de autenticidade e facilitam o engano e a manipulação em ecossistemas digitais", de modo que "a irrupção da IA em nossas vidas nos expõe como nunca à falsidade e ao engano, além de nos desnudar diante de um cenário mundial virtual, nos torna mais vulneráveis e manipuláveis"^[7].

38. Conforme anteriormente exposto pela União, tal cenário de proliferação de desinformação patrocinada em formato de publicidade **coloca em risco a credibilidade de relevantes ações e programas governamentais, além de malferir a reputação de agentes**

públicos, os quais têm suas imagens expostas e manipuladas sem que exista mecanismo hábil, até o momento, para impedir a sua rápida propagação.

39. Nestes termos, as conclusões apresentadas no estudo do NetLab são no sentido de que *"a coexistência de anúncios fraudulentos e anúncios autênticos do Governo Federal gera um cenário confuso e prejudicial, minando a capacidade dos usuários de discernir informações falsas de verdadeiras, além de impactar negativamente na reputação e na confiança na administração pública".*

40. Sendo assim, é salutar que esse Egrégio Supremo Tribunal Federal, ao formular a tese de Repercussão Geral no Tema 987, **avalie explicitar que os deveres anexos**, mencionados no "Decálogo contra a violência digital e a desinformação", **também se aplicam aos casos de produção de conteúdo ilícito por meio de inteligência artificial**, seja para a detecção de fraudes que colocam em risco o êxito de políticas públicas e a credibilidade institucional da Administração Pública, seja na prevenção contra o uso indevido da imagem de figuras públicas.

3. CONCLUSÃO

41. Diante do exposto, em consonância com as manifestações anteriormente apresentadas na qualidade de *amicus curiae*, a União **requer** a essa Egrégia Suprema Corte que, ao formular a tese de Repercussão Geral, considere que os deveres anexos também se aplicam aos casos de uso de inteligência artificial, na forma acima delineada.

Brasília, 21 de maio de 2025

JORGE RODRIGO ARAÚJO MESSIAS

Advogado-Geral da União

ISADORA MARIA B. R. CARTAXO DE ARRUDA

Advogada da União

Secretária-Geral de Contencioso

JOÃO BOSCO TEIXEIRA

Advogado da União

Diretor do Departamento de Controle Difuso

RAFAEL FERNANDES DUBRA

Advogado da União

Coordenador de Ajuizamento e Repercussão Geral em Substituição

PEDRO HENRIQUE MORAIS E SILVA

Advogado da União

Notas

1. [^] Art. 4º, inciso III, da Lei Complementar nº 73/1993.
2. [^] Disponível em <<https://netlab.eco.ufrj.br/post/danos-causados-pela-publicidade-enganosa-na-meta>>.
3. [^] Cresce cada vez mais o número de fraudes e golpes financeiros aplicados em detrimento de usuários, normalmente **valendo-se de perfis falsos, de nomes de instituições públicas e privadas e da imagem de pessoas públicas**, a ponto do Diretor-Geral da Polícia Federal falar na existência de um verdadeiro "cangaço digital" (<https://www1.folha.uol.com.br/mercado/2025/03/e-um-cangaco-digital-diz-chefe-da-pf-sobre-golpes-financeiros.shtml>).

4. [^] Especificamente quanto às crianças, adolescentes e idosos no mundo digital: "(...) o admirável mundo digital se não se fizer acompanhar de redes protetivas e regulação de direitos pode levar a inúmeras **consequências sociais, acentuando desigualdades em especial aos mais vulneráveis** neste processo. Entre os muitos vulneráveis nossos olhares atentos (...) dizem respeito às **crianças, adolescentes e idosos** que são chamados a participarem de conexões virtuais para o exercício de inúmeras atividades cotidianas a depender de inclusões digitais." (SCALQUETTE, Ana C.; VANZOLINI, Patrícia; ROCHA, Renata da; et al. *Whats Up?: Desafios ao Direito: desafios ao direito, inteligência artificial, uso de dados pessoais, Covid-19, direito à saúde, crianças, adolescentes e idosos no mundo digital, biotecnologia e bioética*. São Paulo: Grupo Almedina, 2022. E-book. p.86. ISBN 9786556277110. Disponível em: <https://app.minhabiblioteca.com.br/reader/books/9786556277110/>. Acesso em: 03 mai. 2025).
5. [^] Os denominados celeb-bait - golpes que, não raro, se valem do emprego de inteligência artificial para ludibriar vítimas a partir da imagem de pessoas famosas - já têm sido objeto de providências pela Meta em outras jurisdições. Tal fato indica a viabilidade técnica de se exigir mecanismos de detecção e remoção de anúncios indevidos, inclusive com uso de IA, de modo que sua adoção deve ser viabilizada para as figuras políticas/públicas que assim desejaram, em respeito à legislação brasileira de proteção de dados (<https://oglobo.globo.com/economia/tecnologia/noticia/2025/03/06/meta-usa-reconhecimento-facial-para-detectar-anuncios-falsos-com-celebridades.ghtml>).
6. [^] A notícia e a íntegra da petição inicial podem ser acessadas através do seguinte sítio eletrônico: <https://www.gov.br/agu/pt-br/comunicacao/noticias/agu-move-acao-contra-meta-para-coibir-golpes-que-usam-simbolos-de-governo-e-imagens-manipuladas-nas-redes>.
7. [^] PINTO, Rodrigo Alexandre L.; NOGUEIRA, Jozelia. *Inteligência Artificial e Desafios Jurídicos: Limites Éticos e Legais*. São Paulo: Almedina, 2023. E-book. p.308. ISBN 9786556279268. Disponível em: <https://app.minhabiblioteca.com.br/reader/books/9786556279268/>. Acesso em: 29 abr. 2025.



Documento assinado eletronicamente por JORGE RODRIGO ARAÚJO MESSIAS, com certificado A1 institucional (*.AGU.GOV.BR), de acordo com os normativos legais aplicáveis. A conferência da autenticidade do documento está disponível com o código 2234448419 e chave de acesso 524ec274 no endereço eletrônico <https://sapiens.agu.gov.br>. Informações adicionais: Signatário (a): JORGE RODRIGO ARAÚJO MESSIAS, com certificado A1 institucional (*.AGU.GOV.BR). Data e Hora: 21-05-2025 15:55. Número de Série: 24688056426646610828629120681. Emissor: Autoridade Certificadora do SERPRO Final SSL.



Documento assinado eletronicamente por ISADORA MARIA BELEM ROCHA CARTAXO DE ARRUDA, com certificado A1 institucional (*.AGU.GOV.BR), de acordo com os normativos legais aplicáveis. A conferência da autenticidade do documento está disponível com o código 2234448419 e chave de acesso 524ec274 no endereço eletrônico <https://sapiens.agu.gov.br>. Informações adicionais: Signatário (a): ISADORA MARIA BELEM ROCHA CARTAXO DE ARRUDA, com certificado A1 institucional (*.AGU.GOV.BR). Data e Hora: 21-05-2025 11:11. Número de Série: 24688056426646610828629120681. Emissor: Autoridade Certificadora do SERPRO Final SSL.