

## Dierle Nunes: Regulação da IA e uso de técnicas subliminares

Tramita no Senado o Projeto de Lei (PL) 2.338 que pretende regular o uso da inteligência artificial (IA) no Brasil [1] inspirado em vários aspectos na normativa europeia, buscando compatibilizar uma abordagem baseada em direitos (*rights-based approach*) com uma abordagem baseada no risco (



Isso significa, de um lado, uma preocupação com os direitos

e premissas éticas aplicáveis, como *fairness*, responsabilidade, transparência e ética (acrônimo *Fate* em inglês) em IA, que busca dimensionar o fato de que muitos modelos de IA, a depender do conjunto de dados que o alimenta (*dataset*) ou das escolhas de arquitetura e abordagem, podem gerar respostas enviesadas (discriminação algorítmica), injustas ou antiéticas.

Nesse aspecto, dentro dos limites que uma norma pode gerar de *enforcement* no campo da tecnologia, tenta se promover controles, mas ainda carecendo de uma visão adequada da diversidade geocultural [3], na medida em que ainda somos quase totalmente dependentes de modelos de IA gestados pelos *big players* e de *datasets* pouco adaptados para nossa diversidade cultural.

No entanto, de outro lado, a proposta legislativa, traz uma outra abordagem. Na medida em que a IA se tornou uma tecnologia de propósito geral, como a eletricidade, ela perpassa todos os aspectos das nossas vidas e oferta riscos variáveis a depender de seus usos, trazendo grandes benefícios, mas impondo potenciais danos.

Gera-se a conhecida abordagem embasada em riscos (categorização de riscos — artigos 13 a 18 do PL) [4], com destaque para os modelos de risco excessivo e de alto risco.

Pontue-se que a administração da justiça é considerada um campo de modelagem de IA de alto risco, no artigo 17, VII [5].



Todavia, o PL não deixa claro que, em especial, após o advento do uso mais corrente de *modelos de fundação* [6], como v.g. os GPTs e o Bert, que servem de base para as IAs generativas (*GEN AIs* como *ChatGPT*, *Midjourney*, *Gamma.app* etc) com propósitos bem diversificados, a simples classificação de riscos não será suficiente para estabelecer uma governança adequada. Inclusive os eurodeputados estão propondo a necessidade de um resumo dos dados de treinamento desses modelos cobertos pela lei de direitos autorais desses novos modelos [7].

É fundamental reconhecer que até as Inteligências Artificiais analíticas, mais tradicionais, apresentam riscos. Foi, de fato, a partir delas que emergiram inquietações relacionadas a discriminações algorítmicas, como evidenciado pelo conhecido caso do *Compas* [8].

Destaque-se que, um dos temas mais sensíveis nas propostas de regulamentação refere-se à adoção de cláusulas gerais e conceitos indeterminados. Um deles é o termo "técnicas subliminares" que se refere a instrumentos que visam coibir práticas reconhecidas que buscam influenciar indivíduos, induzindo a certos comportamentos. Essa influência varia desde estímulos simples para compras até interferências significativas em eleições, afetando a própria sustentação da democracia. Portanto, a preocupação com o uso de sistemas de IA para manipulação já é antiga.

O que muitos desconhecem é que tais regulações não podem se limitar ao controle das tecnologias de IA, uma vez que tal indução de comportamentos passa pelo uso do design comportamental (e seus padrões obscuros — *dark patterns*) [9] que se vale da utilização do conhecimento dos processos decisórios humanos, que trabalham com racionalidade limitada.

Em verdade, a virada tecnológica no direito [10] importa uma análise das questões jurídicas mediante uma visão interdisciplinar sobre os efeitos da tecnologias da informação e comunicação, com destaque para IA, somado ao do uso do design comportamental (e persuasivo), mediante o emprego da psicologia cognitiva comportamental [11] e da neurociência.

Nesses termos, as regulações propostas são limitadas por não entender a extensão do problema e ao usar termos amplos e indeterminados, terão um grande desafio de aplicação (*enforcement*)

Frequentemente, emprega-se o termo "técnicas subliminares" para referir-se a aplicações de IA consideradas de risco excessivo (proibidos).

O projeto de Lei da IA da União Europeia as define e as proíbe no seu artigo 5º, 1, "a" nos seguintes termos: "*estão proibidas as seguintes práticas de inteligência artificial: a) a colocação no mercado, a colocação em serviço ou a utilização de um sistema de IA que empregue técnicas subliminares que contornem a consciência de uma pessoa para distorcer substancialmente o seu comportamento de uma forma que cause ou seja suscetível de causar danos físicos ou psicológicos a essa ou a outra pessoa*" [12].



Inspirado no *AIA Act* o artigo 14 do projeto de lei 2338/23 brasileiro veda o uso de técnicas subliminares, as definindo como de risco excessivo, do seguinte modo: *"são vedadas a implementação e o uso de sistemas de inteligência artificial: I – que empreguem técnicas subliminares que tenham por objetivo ou por efeito induzir a pessoa natural a se comportar de forma prejudicial ou perigosa à sua saúde ou segurança ou contra os fundamentos desta Lei"*.

Nesses termos, além da insuficiência da norma em termos de *enforcement* não se traz uma conceituação adequada do que se trata.

Tradicionalmente, "subliminar" refere-se a estímulos que escapam à percepção consciente, mas ainda influenciam o comportamento, como imagens exibidas por um tempo extremamente curto. Definir técnicas subliminares de forma restrita, como simples estímulos sensoriais imperceptíveis, é problemático porque muitas formas potencialmente manipuladoras de influência não se enquadram nessa categoria [\[13\]](#).

As técnicas de manipulação problemáticas muitas vezes não se escondem apenas no nível sensorial, mas também na intenção, no método e no efeito da influência. Por exemplo, os usuários podem não saber que o conteúdo online que veem é adaptado especificamente para eles (microdirecionamento) [\[14\]](#), nem compreender os métodos ou os efeitos dessa adaptação [\[15\]](#).

Uma proposta de definição para técnicas subliminares poderia identificá-las como aquelas que buscam influenciar o comportamento de uma pessoa sem que ela esteja ciente da tentativa de influência, de como essa influência opera, ou dos seus efeitos nas decisões e crenças da pessoa. Isso deixa claro o enfraquecimento de sua autonomia privada e de seu consentimento informado. Esta definição inclui a maioria das formas preocupantes de influência impulsionada pela IA e fornece uma estrutura mais aplicável.

Para que a regulação seja eficaz, é crucial que "técnicas subliminares" sejam definidas de maneira abrangente, abordando todas as formas potencialmente manipuladoras de influência, para garantir a proteção dos usuários sem sobrecarregar os prestadores de serviços.

Em recente estudo sobre o tema no *AI Act*, Bermúdez *et al* propõem uma definição. Segundo eles:

*a escolha de técnicas subliminares como alvo da proibição pode levar a interpretações da proibição que são demasiado restritas para abordar os problemas mais graves e formas generalizadas de manipulação problemática, ou a uma leitura que é tão ampla que impõe encargos proibitivos a qualquer organização que pretenda desenvolver ou utilizar sistemas de IA. Para evitar ambos os extremos, propomos uma definição que identifica as técnicas subliminares como as técnicas de influência em que o agente influenciado permanece previsivelmente inconsciente do estímulo influenciador, da forma como as técnicas operam ou dos seus efeitos nos seus valores, crenças e decisões. Esta definição baseia-se em opiniões amplamente compartilhadas sobre a autonomia pessoal, oferece uma interpretação coerente do texto jurídico e fornece orientação ética para a prática de desenvolvimento tecnológico* [\[16\]](#).

Uma abordagem que talvez mereça ser somada a essa sejam os comandos propostos de regulação de design (como no projeto de *Detour Act* americano). Se aprovada, a legislação autorizará a Comissão Federal de Comércio dos EUA a avaliar se uma ação ou conduta é injusta ou enganosa ao "comprometer



ou ameaçar a liberdade de escolha do usuário, ao solicitar consentimento ou dados, ou ao incentivar o uso obsessivo por crianças". O texto define "*experimento ou pesquisa comportamental ou psicológica*" como "*análise das ações visíveis ou dos processos mentais, percebidos através do comportamento, incluindo as interações individuais e coletivas*" [17].

Será proibido, caso aprovado, para grandes plataformas online [18]:

- 1) Alterar ou configurar interfaces para comprometer a livre escolha ou autonomia do usuário ao solicitar consentimento ou dados;
- 2) Segmentar os usuários em grupos específicos para pesquisas comportamentais ou psicológicas sem o consentimento expresso de cada usuário;
- 3) Modificar interfaces em plataformas destinadas a crianças com o propósito de estimular uso excessivo, incluindo vídeos que são reproduzidos automaticamente sem a autorização do usuário.

Além disso, a proposta impõe obrigações aos grandes provedores, reforçando a necessidade de obter consentimento claro e informado dos usuários.

Destaca-se, finalmente, como pressuposto desse debate, que já se defende a imperatividade de reconhecer novos direitos fundamentais, os neurodireitos, que salvaguardem a liberdade cognitiva e os direitos concernentes à privacidade mental, integridade mental e continuidade psicológica [19], para defesa e preservação da autonomia privada [20], conforme já previsto na Constituição Chilena [21].

Vê-se, assim, que a regulação da IA e de seus impactos em ambientes decisórios precisa ser realizada com uma abordagem essencialmente interdisciplinar e ampla, de modo a promover um controle sofisticado para uma tecnologia em avanço constante, sem gerar uma barreira à inovação.

[1] <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>

[2] Em agosto foi criada uma Comissão Temporária Interna para discussão o destina-se a examinar, no prazo de 120 dias, os projetos concernentes ao relatório final aprovado pela Comissão de Juristas responsável por subsidiar a elaboração de substitutivo sobre Inteligência Artificial no Brasil (Cjusbia), criada pelo Ato do Presidente do Senado Federal nº 4, de 2022:

<https://legis.senado.leg.br/comissoes/comissao?codcol=2629>

[3] PRABHAKARAN, Vinodkumar; MITCHELL, Margaret; GEBRU, Timnit; GABRIEL Iason. *A Human Rights-Based Approach to Responsible AI. 2022 ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization or (EAAMO '22)*. Acessível em:



---

<https://arxiv.org/abs/2210.02667>

[4] <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>

[5] Artigo 17...VII – administração da justiça, incluindo sistemas que auxiliem autoridades judiciárias na investigação dos fatos e na aplicação da lei.

[6] Os modelos de fundação *"são treinados em dados amplos em grande escala e são adaptáveis a uma ampla gama de tarefas posteriores"*. BOMMASANI, Rishi et al. *On the Opportunities and Risks of Foundation Models*. Stanford Institute for Human-Centered Artificial Intelligence (HAI), 2022. <https://arxiv.org/abs/2108.07258>

[7] <https://rr.sapo.pt/noticia/politica/2023/06/13/eurodeputados-e-o-chatgpt-inteligencia-artificial-e-comparavel-a-energia-nuclear/335031/>

[8] NUNES, Dierle; MARQUES, Ana Luiza Pinto Coelho. Inteligência artificial e direito processual: vieses algorítmicos e os riscos de atribuição de função decisória às máquinas. *Revista de Processo*, v.285, nov./2018

[9] Como informa Waldman: *"Os designers usam padrões obscuros para esconder, enganar e incitar os usuários a revelar. Eles confundem os usuários fazendo perguntas de maneiras que não especialistas não conseguem entender, ofuscam ocultando elementos de interface que podem ajudar os usuários a proteger sua privacidade, exigem registro e divulgações associadas para acessar a funcionalidade e ocultam comportamento malicioso no abismo das políticas de privacidade legais. Os padrões escuros também tornam a divulgação "irresistível" ao conectar o compartilhamento de informações aos benefícios do aplicativo"*. WALDMAN, Ari. *Cognitive Biases, Dark Patterns, and the 'Privacy Paradox'*. *31 Current Issues in Psychology 2020*. Nos EUA o projeto de Detour Act (S. 3330) em tramitação no Congresso Americano, visa *"proibir o uso de práticas exploratórias e enganosas por grandes operadoras online e promover o bem-estar do consumidor no uso de pesquisas comportamentais por esses provedores"* abrangendo todo serviço on-line, que não seja de acesso à Internet, disponibilizado pela Web como as redes sociais, mecanismos de pesquisa e de e-mail, viabilizando maior transparência no ambiente on-line. Fica evidente na proposta busca combater os padrões obscuros no design (*dark patterns*) que estruturam uma interface de usuário cuidadosamente elaborada para induzir os usuários a comportamentos não desejados.

<https://www.congress.gov/bill/117th-congress/senate-bill/3330/text>

[10] NUNES, Dierle. Virada tecnológica no direito processual: fusão de conhecimentos para geração de uma nova justiça centrada no ser humano. *RePro*. V. 344. 2023.



---

[11] NUNES, Dierle et al. *Desconfiando da imparcialidade dos sujeitos do processo*. São Paulo: Juspodivm, 2022.

[12] <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>

[13] <https://www.euractiv.com/section/artificial-intelligence/opinion/the-ai-act-needs-a-practical-definition-of-subliminal-techniques/>

[14] Microtargeting é a utilização de dados online para adaptar mensagens publicitárias e de marketing personalizado a indivíduos, com base na identificação das vulnerabilidades pessoais dos destinatários. Cf. LORENZ-SPREEN, Philipp; et al. *Boosting people's ability to detect microtargeted advertising*. *Scientific Reports*. 2021. Acessível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8324838/>

[15] Idem.

[16] BERMÚDEZ, Juan Pablo. *What Is a Subliminal Technique? An Ethical Perspective on AI-Driven Influence*. 2023. <https://philpapers.org/rec/BERWIA-9>

[17] <https://www.congress.gov/bill/117th-congress/senate-bill/3330/text>

[18] Entendido o "termo 'grande operador online' como qualquer pessoa que (A) fornece um serviço online; (B) tenha mais de 100.000.000 usuários autenticados de um serviço online em qualquer período de 30 dias; e (C) está sujeita à jurisdição da Comissão sob a Lei da Comissão Federal de Comércio (15 USC 41 et seq.)".

[19] YUSTE, Rafael; GOERING, Sara; et. al. *Four ethical priorities for neurotechnologies and AI*. *Nature*, Londres; 2017.

[20] NUNES, D. *Autonomia privada e novas tecnologias: rumo aos neurodireitos*. <https://spotify.link/h8GRP753aDb>

[21] *Chile: LEY 21383/21 modifica la carta fundamental, para establecer el desarrollo científico y tecnológico al servicio de las personas*. <https://www.bcn.cl/leychile/navegar?idNorma=1166983&idParte=10278855&idVersion=2021-10-25>

## Meta Fields