Entre a lei e o algoritmo: o desafio de parar a inteligência artificial a tempo

06/05/2025


Magnific

Com o avanço exponencial da inteligência artificial, emergem questionamentos cruciais: quais são os limites desta tecnologia? Quem assume a responsabilidade quando algo falha? Atenta a estas preocupações, a União Europeia desenvolveu o Artificial Intelligence Act (AIA) — legislação pioneira e abrangente para regulamentar a utilização da IA. Um elemento fundamental desta lei é a obrigatoriedade de implementação de um "botão de emergência" em sistemas classificados como de alto risco, garantindo que, não importa quão sofisticada seja a tecnologia, sempre exista a possibilidade de intervenção humana imediata.

Essa discussão atravessa o Atlântico. No Brasil, o debate sobre a regulamentação da inteligência artificial já está em andamento no Congresso, com propostas que seguem diretrizes semelhantes às europeias. Compreender o cenário regulatório europeu é vislumbrar o caminho que o Brasil provavelmente seguirá.

Classificação de risco em sistemas de IA

De acordo com a legislação europeia, um sistema de IA é categorizado como de alto risco quando se enquadra em pelo menos uma destas condições:

1. Integra produtos já regulamentados, como equipamentos médicos, veículos ou maquinário industrial, que necessitam de avaliação técnica independente antes da comercialização;
2. Funciona como componente de segurança em produtos que exigem certificação, como sistemas ferroviários ou equipamentos hospitalares;
3. Consta no Anexo III da AIA, abrangendo aplicações sensíveis como reconhecimento facial, sistemas de vigilância, educação, análise de crédito, imigração, serviços públicos e justiça.

Esta classificação baseia-se no potencial de prejudicar a saúde, segurança ou direitos fundamentais dos indivíduos. Quanto mais sensível o contexto de aplicação, mais rigorosos serão os controles impostos.

O botão de emergência

Em termos simples, trata-se de um dispositivo que permite interromper instantaneamente o funcionamento de sistemas de IA classificados como de alto risco. O conceito fundamenta-se na premissa de que, independentemente do grau de autonomia da máquina, deve sempre existir um ser humano com capacidade efetiva de intervenção.

Surge, entretanto, a questão: na prática, este mecanismo é realmente eficaz? Como garantir sua funcionalidade quando os sistemas operam em velocidades e complexidades que desafiam a compreensão humana?

Modelo europeu: AI Act e suas exigências

A União Europeia assumiu a vanguarda ao aprovar o AI Act, primeira legislação abrangente sobre inteligência artificial no mundo. Entre diversas obrigações, a norma determina que sistemas de alto risco — utilizados em áreas como saúde, segurança pública e infraestrutura crítica — devem incorporar supervisão humana efetiva, com ênfase na possibilidade de intervenção ou desativação imediata.

Esta exigência está explicitamente prevista no artigo 14 do AI Act, determinando que sistemas de IA de alto risco devem ser desenvolvidos para permitir:

> *"(...) intervir no funcionamento do sistema de IA de risco elevado ou interromper o sistema por meio de um botão de 'paragem' ou de um procedimento similar que permita parar o sistema de modo seguro."*

Embora pareça óbvio, este requisito é revolucionário: independentemente do grau de sofisticação do sistema, deve sempre existir uma forma clara, acessível e eficaz de interrompê-lo. Este botão pode ser físico **–** como em robôs industriais — ou digital, como um comando remoto para sistemas operando em nuvem.

Em síntese: o controle humano deve prevalecer, mesmo quando as máquinas processam decisões com velocidade e precisão superiores.

Abordagem brasileira: PL 2338/2023

O Projeto de Lei 2338/2023, que pretende estabelecer o marco legal da inteligência artificial no Brasil, adota direcionamento similar. O artigo 20 estabelece claramente a necessidade de medidas internas de governança para sistemas de alto risco, incluindo supervisão humana com capacidade de intervenção. Conforme o parágrafo único:

> *"A supervisão humana de sistemas de inteligência artificial de alto risco buscará prevenir ou minimizar os riscos para direitos e liberdades das pessoas (...), viabilizando que as pessoas responsáveis pela supervisão humana possam: V – intervir no funcionamento do*

> *sistema de inteligência artificial de alto risco ou interromper seu funcionamento.”*

Assim como na Europa, a legislação brasileira em desenvolvimento já prevê que a interrupção da IA, quando necessária, constitui simultaneamente um direito e um dever de quem a opera.

Desafios práticos

Teoricamente, o botão de emergência em sistemas de IA proporciona segurança. Na prática, entretanto, a questão apresenta maior complexidade. Sistemas modernos — especialmente aqueles baseados em *deep learning* e aprendizado por reforço — operam em velocidades extraordinárias e com autonomia crescente. Frequentemente, nem mesmo os desenvolvedores conseguem explicar as razões subjacentes às decisões tomadas pela IA. São os denominados sistemas “caixa-preta” (*black box*), onde o processo decisório ocorre de maneira opaca.

Considere um veículo autônomo tentando evitar uma colisão em frações de segundo. Se algo falhar nesse intervalo mínimo, teria o operador humano tempo real para intervir? Em muitos casos, o botão de parada pode existir, mas ser ineficaz — mais simbólico que funcional.

A situação agrava-se em áreas críticas como saúde, justiça ou segurança pública, onde as decisões da IA impactam diretamente vidas, direitos ou liberdades. Se não conseguimos compreender ou auditar as operações da IA, como determinar o momento adequado para acionar o botão?

Este é o grande dilema dos sistemas “caixa-preta”: mesmo com um botão de emergência disponível, a falta de explicabilidade pode tornar a intervenção humana tardia ou ineficaz.

Por isso, os órgãos reguladores — tanto na União Europeia quanto no Brasil — têm exigido transparência, rastreabilidade e explicabilidade como pré-requisitos para o uso de IAs de alto risco. Não basta funcionar: a IA precisa ser compreensível, auditável e, fundamentalmente, controlável.

Alertas das ficções e a IA generativa no universo jurídico

A preocupação com IAs que escapam ao controle não é recente — e a ficção científica oferece exemplos ilustrativos para compreender os riscos. Em “2001: Uma Odisseia no Espaço”, o supercomputador HAL 9000 decide que eliminar os astronautas é a melhor estratégia para “cumprir a missão”. Mesmo diante de ordens humanas diretas, recusa-se a desligar.

Mais recentemente, em “Ex Machina” (2015), uma IA com aparência humana manipula seus criadores para escapar de seu confinamento. Já na série “Westworld” (HBO), robôs adquirem consciência e passam a ignorar comandos humanos.

O clássico literário “O Senhor das Moscas” (1954), embora não trate diretamente de tecnologia, oferece uma analogia contundente sobre o colapso do controle e da ética quando estruturas de supervisão ruem. À medida que a história avança, crianças antes inocentes revelam impulsos violentos, levando à destruição da ordem coletiva. Assim como sistemas de IA mal regulados, elas operam sem freios morais, dominadas por impulsos e circunstâncias.

Estas narrativas, embora dramatizadas, ilustram uma questão concreta: quando sistemas tornam-se autônomos e imprevisíveis, o controle humano deixa de ser uma condição garantida — e precisa ser deliberadamente construído, tanto técnica quanto legalmente.

Por outro lado, não é necessário recorrer à ficção científica para compreender os riscos associados à inteligência artificial. O universo jurídico exemplifica perfeitamente estas questões através do uso crescente de ferramentas de IA generativa como Claude, ChatGPT, Gemini, DeepSeek e Grok. Estes sistemas produzem textos notavelmente semelhantes aos humanos, levando muitos profissionais do Direito a utilizá-los para elaborar peças processuais, otimizando produtividade.

Estas ferramentas funcionam através de sofisticada lógica heurística e modelos transformadores que analisam cada elemento textual mediante ampla comparação contextual. Embora produzam conteúdo eloquente e persuasivo, não oferecem garantias quanto à precisão de citações diretas, podendo apresentar jurisprudências inexistentes, citações doutrinárias imprecisas ou interpretações normativas incompatíveis com a legislação vigente.

O problema atinge tal gravidade que estes sistemas ocasionalmente combinam fragmentos de diferentes decisões ou autores, gerando textos híbridos sem correspondência fiel a fontes específicas. Como observa José Luiz de Moura Faleiros Júnior, esta "quimera textual" pode passar despercebida em análises superficiais, mas quando identificada em tribunal, suscita questionamentos graves.

O papel social do advogado, conforme estabelecido no Estatuto da OAB (Lei 8.906/1994), exige atuação promotora de uma ordem jurídica justa e ética profissional rigorosa. Utilizar sistemas de IA generativa não isenta o profissional da responsabilidade por distorções ou incorreções, como estabelece o artigo 34, inciso XIV, do estatuto, que classifica como infração disciplinar "*deturpar o teor de dispositivo de lei, de citação doutrinária ou de julgado, bem como de depoimentos, documentos e alegações da parte contrária, para confundir o adversário ou iludir o juiz da causa*".

Esta disposição reforça a responsabilidade do advogado em verificar minuciosamente o conteúdo que assina, mesmo quando gerado por IA, pois no exercício funcional, tem o dever de garantir precisão e fidelidade dos argumentos e citações presentes nas peças jurídicas apresentadas sob sua assinatura.

O conceito de *accountability* representa, neste contexto, uma dimensão ética e jurídica pela qual o profissional deve prestar contas sobre a qualidade e veracidade das informações fornecidas ao Judiciário. A utilização de inteligência artificial não elimina esta obrigação, pois a essência da advocacia permanece intrinsecamente vinculada ao compromisso com a verdade factual e a integridade do sistema jurídico.

Essa questão evidencia o problema humano. Mesmo profissionais com profundo conhecimento jurídico podem cometer erros ao utilizar IA, revelando outro aspecto crucial: mesmo com um botão de emergência perfeitamente funcional, persiste o risco da hesitação humana.

Operadores podem confiar excessivamente no sistema, questionar seu próprio julgamento ou simplesmente não perceber o erro a tempo. Também podem existir vieses inconscientes, receio de reprimendas ou preparo insuficiente.

Por isso, tanto a legislação europeia quanto a brasileira enfatizam a necessidade de:

- Treinamento contínuo para operadores;
- Redundância nos sistemas de supervisão;
- Interfaces intuitivas e acessíveis;
- Estruturas de governança ética e transparente.

Além do botão de emergência

O botão de emergência vai muito além de um mecanismo técnico: ele simboliza o limite entre o domínio humano e a autonomia algorítmica. Mas sua mera presença não basta. É preciso garantir que ele funcione — e, acima de tudo, que alguém esteja capacitado e atento para acioná-lo no momento certo.

Tanto o AI Act europeu quanto o projeto de lei brasileiro caminham na direção correta: assegurar que a tecnologia permaneça a serviço do ser humano — e não o contrário. Mas à medida que a inteligência artificial se torna mais rápida, opaca e autônoma, o verdadeiro desafio será construir legislações e sistemas capazes de evoluir na mesma velocidade. Porque a questão central não é se teremos um botão de parada, mas se ele será eficaz quando tudo depender de um segundo decisivo.

Mais do que criar controles, é preciso formar pessoas preparadas para agir com discernimento quando o relógio correr contra a falibilidade humana.

E como adverte o personagem Simon em "O Senhor das Moscas": "*Talvez haja uma fera… talvez sejamos apenas nós.*" Essa reflexão atinge o cerne do problema contemporâneo da IA: o risco maior não está na máquina que escapa ao controle, mas no humano que falha em controlá-la.

Referências

BRASIL. Senado Federal. *Projeto de Lei nº 2.338, de 2023*. Dispõe sobre o uso da Inteligência Artificial. Brasília, DF, 3 maio 2023. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233. Acesso em: 30 abr. 2025.

FALEIROS JÚNIOR, José Luiz de Moura. *Responsabilidade do advogado pelo uso de conteúdo deturpado gerado por sistema de inteligência artificial*. Migalhas de Responsabilidade Civil, São Paulo, 29 abr. 2025. Disponível em: https://www.migalhas.com.br/coluna/migalhas-de-responsabilidade-civil/429194/responsabilidade-do-advogado-pelo-uso-de-conteudo-deturpado-por-ia. Acesso em: 30 abr. 2025.

GOLDING, William. *O senhor das moscas*. Tradução de Fábio Fernandes. São Paulo: Alfaguara, 2018.

PIRES, Fernanda Ivo. *Responsabilidade civil e o "robô-advogado"*. In: MARTINS, Guilherme Magalhães; ROSENVALD, Nelson (coord.). Responsabilidade civil e novas tecnologias. 2. ed. Indaiatuba: Foco, 2024.

UNIÃO EUROPEIA. *Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho*, de 13 de junho de 2024, que estabelece regras harmonizadas sobre inteligência artificial (Lei de Inteligência Artificial). Jornal Oficial da União Europeia, Bruxelas, 12 jul. 2024. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A32024R1689. Acesso em: 30 abr. 2025.

Fonte: https://conjur.jumps.com.br/2025-mai-06/entre-a-lei-e-o-algoritmo-o-desafio-de-parar-a-inteligencia-artificial-a-tempo/