Visão do Direito

Daniel Becker
Sócio das áreas de contencioso e arbitragem, proteção de dados e inteligência artificial do BBL Advogados, membro das comissões de 5G e assuntos legislativos da OAB/RJ

Regulação da IA: o Brasil precisa evitar os erros do passado

A recente instalação da comissão especial na Câmara dos Deputados para analisar o projeto de regulamentação da inteligência artificial (IA) no Brasil marca um momento crucial. O texto, aprovado pelo Senado em 2023, propõe a criação do Sistema Nacional de Regulação e Governança de IA (SIA), sob coordenação da ANPD, e abre caminho para um debate necessário. No entanto, é preciso aprender com os equívocos recentes — especialmente os do PL das Fake News — para não repetir erros que possam comprometer a eficácia da futura lei.

Um dos principais riscos nesse processo é a tentativa de fundir a regulação da IA com o controle das plataformas digitais, como ocorreu no PL 2630. Naquele caso, a legítima preocupação com a desinformação se perdeu em meio a um debate paralelo — e mal conduzido — sobre o poder econômico das big techs. São temas distintos, com fundamentos normativos diferentes, e misturá-los gera soluções frágeis, que não resolvem nenhum dos problemas e ainda criam insegurança jurídica.

A IA não deve ser tratada como um problema a ser contido, mas como uma ferramenta cujos riscos dependem de seu uso. O perigo não está na tecnologia em si, mas em como ela é projetada e aplicada. Regular a IA com base em seus impactos concretos — e não em abstrações — é o caminho mais seguro para evitar excessos que possam sufocar a inovação.

Em vez de uma regulação engessada, precisamos de uma "autorregulação regulada", em que o setor privado proponha protocolos e medidas de mitigação, sob supervisão do Estado. Inovação responsável exige regulação inteligente, não rigidez. Essa abordagem já encontra eco até mesmo na União Europeia, que, após aprovar o AI Act, prepara-se para lançar um Código de Boas Práticas em parceria com o GPAI (Global Partnership on AI) — um reconhecimento tácito de que leis rígidas, sozinhas, não dão conta da velocidade da evolução tecnológica.

O Brasil tem a oportunidade de adotar um marco regulatório moderno, que proteja direitos fundamentais sem engessar o desenvolvimento tecnológico. Para isso, é essencial focar nos usos e impactos da IA, não na tecnologia em abstrato, além de evitar a tentação de regular plataformas e IA no mesmo pacote, sob o risco de repetir os erros do PL das Fake News. Ou seja, é preciso adotar modelos flexíveis de governança, combinando regulação estatal com diretrizes setoriais.

Enquanto o Congresso avança na discussão, é preciso lembrar que uma regulação mal desenhada pode ser pior do que nenhuma regulação. O Brasil não pode perder a chance de construir um marco jurídico que estimule a inovação, garanta segurança jurídica e proteja a sociedade — mas, para isso, deve evitar os atalhos que já se mostraram falhos no passado.

A inteligência artificial é uma revolução em curso. Cabe ao Brasil decidir se será um mero espectador ou um protagonista dessa transformação. A escolha começa agora.

Visão do Direito

Luis Fernando Prado
Sócio no escritório Prado Vidigal, membro do conselho consultivo da Associação Brasileira de Inteligência Artificial (Abria) e LLM em direito digital e sociedade da informação pela Universidade de Barcelona

Inteligência artificial e Large Language Models: tensões para a proteção de dados

É muito provável que, nos últimos anos, você tenha utilizado assistentes virtuais, chatbots, ferramentas de redação e edição de textos, aplicativos de tradução e diversas outras soluções rotuladas como inteligência artificial (IA), que facilitam o nosso dia a dia. Por trás do desempenho de muitas dessas ferramentas estão os Large Language Models (LLMs), modelos treinados com base em grandes volumes de dados textuais para gerar linguagem humana de forma avançada, sendo capazes de realizar variadas tarefas linguísticas, como tradução, resumo de textos, resposta a perguntas, análises, correlações de dados e até criação de conteúdo.

Durante interações com sistemas de IA que contam com um LLM entre seus componentes, é natural que os inputs (informações e instruções fornecidas no início da interação) e os outputs (resultados gerados pela ferramenta) possam envolver dados pessoais — especialmente, no caso dos outputs, quando informações relacionadas a pessoas físicas são solicitadas na entrada. Diante desse cenário, surgem questionamentos legítimos: como o modelo consegue retornar dados pessoais? Esses dados já estavam previamente armazenados pelos componentes do sistema?

Compreender como os elementos textuais são processados e armazenados em LLMs é essencial para responder a essas perguntas. Por isso, a Autoridade de Proteção de Dados e Liberdade da Informação de Hamburgo publicou um material de discussão que avalia aspectos técnicos do funcionamento dos LLMs à luz das regras e obrigações existentes no campo da proteção de dados pessoais. Nesse sentido, a autoridade de Hamburgo destaca que um aspecto técnico fundamental é o processo de tokenização do texto de entrada, que, em síntese, significa que todos os textos são divididos em pedaços menores antes de serem inseridos no modelo.

Assim, é importante notar que palavras longas ou frases não são incorporadas diretamente ao modelo. Por exemplo, os elementos textuais "Luis Fernando Prado" não estão armazenados no modelo, que contém apenas tokens individuais, como "Lui", "Pra" etc. A partir desses fragmentos, o processo de treinamento envolve o aprendizado de correlações posicionando os tokens uns em relação aos outros, ou seja, atribuindo-lhes pesos de probabilidade com base na coerência percebida para o usuário final.

Na prática, isso significa que eventuais dados pessoais são convertidos em representações matemáticas abstratas, que capturam padrões gerais e correlações derivadas dos dados de treinamento como um todo. Por isso, a autoridade de Hamburgo reconhece que esse processo de abstração resulta na perda de características concretas e referências a indivíduos específicos e, consequentemente, nenhum dado pessoal é armazenado nos LLMs.

Mas então, como explicar o fato de que determinadas ferramentas que incorporam LLMs podem apresentar dados pessoais em suas saídas e resultados? Os textos de treinamento são processados com base em relações contextuais e, por isso, os outputs podem ser semelhantes — ou coincidir parcialmente — com os textos de treinamento. No entanto, a mera presença de informações pessoais plausíveis na saída do modelo não significa que dados pessoais foram "memorizados". Na verdade, os resultados produzidos por LLMs são criações probabilísticas, e não simples reproduções de informações armazenadas e recuperadas de uma base de dados específica, como ocorreria em um documento de texto.

No campo da proteção de dados, compreender adequadamente o funcionamento dos LLMs é essencial para definir os contornos de aplicação das regras e obrigações previstas em legislações e normas existentes. Por exemplo, considerando que nenhum dado pessoal é armazenado por um LLM, os direitos previstos pela Lei Geral de Proteção de Dados (Lei nº 13.709/2018 LGPD) não podem ser exercidos em face do modelo em si, mas apenas em relação às entradas e saídas de um sistema de IA. Mesmo nesse caso, a explicabilidade sobre como os modelos processam informações e geram resultados é inerentemente complexa, uma vez que a linguagem é dinâmica e dependente de contexto — o que gera desafios na previsão de como os LLMs se comportarão nos mais diversos cenários do mundo real.

Reconhecer essas diferenças e particularidades técnicas permite que regulamentações de proteção de dados sejam aplicadas de forma adequada a cada contexto tecnológico. A autoridade de proteção de dados de Hamburgo já deu um primeiro passo nessa direção, demonstrando que é necessário compreender aspectos técnicos e operacionais para, então, aplicar corretamente os controles existentes nas regulamentações.

No Brasil, é importante lembrar que a regulação da IA está em discussão na esfera legislativa e também vem sendo objeto de análise por autoridades como a Autoridade Nacional de Proteção de Dados (ANPD). Por isso, no atual contexto de rápida evolução tecnológica e de amadurecimento do debate regulatório sobre IA, é fundamental garantir o equilíbrio entre as normas de proteção de dados já existentes e o desenvolvimento ético e seguro de sistemas de IA — incluindo seus componentes, como é o caso dos LLMs.

DIÁRIOS ASSOCIADOS | **Diretora de Redação:** Ana Dubeux; **Edição:** Ana Maria Campos; **Repórter**: Maria Eduarda Lavocat; **Diagramação:** Arthur Filho; **Arte e Ilustrações:** Kleber Sales