# Governança em inteligência artificial: riscos e fatores mínimos de prevenção

02/01/2024

Ainda não se chegou a um consenso sobre a regulamentação da inteligência artificial (IA) no Brasil. Em 2023, tramitava no Congresso Nacional o PL nº 2.338/2023, o qual veio a receber proposta substitutiva datada de 27 de novembro, assinada pelo senador Astronauta Marcos Pontes. Mesmo longe de uma resposta legislativa, há na indústria um consenso a respeito dos riscos inerentes ao desenvolvimento e utilização dessas tecnologias, o que desde já impõe o emprego de medidas de mitigação aos principais riscos e potenciais danos aos indivíduos. Fica, assim, ao encargo dos programas internos de governança em IA a adoção de providências necessárias ao desenvolvimento e à utilização responsáveis das aplicações de IA. Neste ensaio, propõe-se pontuar quais as fórmulas possíveis para o enfrentamento dos principais riscos.

Todas as propostas de regulamentação circulando globalmente apontam o dever de transparência, exigindo que a finalidade dos sistemas de IA seja abertamente informada e as decisões sejam explicáveis a quem seja por elas afetado. O efeito "*black box*" é, assim, um dos principais riscos a se evitar. E, para tanto, os profissionais devem garantir que os seus algoritmos sejam explicáveis e acessíveis. É certo que a explicabilidade é desafiadora em face dos limites que os próprios sistemas impõem e da proteção do segredo de indústria, impedindo uma efetividade plena da norma legal. Assim, dependendo da complexidade dos modelos de IA, especialmente aqueles baseados em aprendizado profundo e o volume de dados utilizados para treinamento, o processo interno de tomada de decisão é de difícil compreensão e consequente explicação, até mesmo para os especialistas.


Mesmo assim, é também certo que o cidadão deve estar informado sempre que as decisões que impactam em sua vida forem decorrentes de sistemas automatizados de IA. Daí ser indispensável privilegiar o dever de transparência, o qual inclui a identificação de sistemas que não são integralmente explicáveis. Por isso, há que estabelecer requisitos mínimos que garantam um nível aceitável de transparência e responsabilidade, incluindo informações sobre a descrição do funcionamento geral algorítmico, indicadores de desempenho e limitações algorítmicas, métricas de precisão, confiabilidade e áreas onde há maior exposição a vieses.

Por tudo isso, o acesso às informações sobre o propósito da coleta e do processamento de dados, e ainda sobre quem sejam o desenvolvedor e os usuários do algoritmo, deve ser garantido ao cidadão. Como consequência, também deve ser disponibilizado um canal de comunicação, permitindo que os cidadãos façam perguntas e recebam mais informações. A cidade de Amsterdã, por exemplo, lançou os "registros de IA abertos", que oferecem uma visão

geral dos sistemas de inteligência artificial existentes e dos algoritmos usados pelo governo municipal. O objetivo central dos registros é “ser aberto e transparente sobre o uso de algoritmos”, explicando como é feita a coleta e o processamento de dados, como algoritmos evitam a discriminação, quais os riscos e salvaguardas, e como a supervisão humana é implementada.

Os deveres anteriores são apenas partes integrantes de um dever maior: o da responsabilidade (*accountability*). Para responder à exigência de *accountability*, devem ser estabelecidos órgãos de supervisão internos e externos, a eles destinando financiamento adequado e recursos humanos qualificados. Note-se que a responsabilidade desses órgãos se estende para além da análise de algoritmos para todos os aspectos do uso de dados, incluindo os meios de coleta de dados, o propósito, o processamento e o armazenamento de dados, e o uso dos resultados (incluindo os usos secundários).  Em geral, as equipes de ética nas empresas de tecnologia respondem por essas medidas.

Na prática, *accountability* exige: indicar quem são os agentes responsáveis (controladores e operadores dos algoritmos); definir procedimentos para colaboração com autoridades competentes ao lidar com conteúdos ilegais, por exemplo, decisões discriminatórias; e estabelecer a responsabilidade dos desenvolvedores pela detecção e mitigação de conteúdos ilegais gerados por suas IA.

Ainda é necessário todo o cuidado para garantir a qualidade dos dados, uma vez que a qualidade das entradas (*inputs*) determinará a qualidade das saídas (*outputs*). Uma estratégia adequada de coleta e qualidade de dados pode mitigar muitos problemas, por isso, monitorar a qualidade e a coleta de dados é fundamental para evitar vieses e aplicações discriminatórias. Isso exigirá que os desenvolvedores de IA sejam transparentes sobre como os modelos de linguagem foram treinados e quais dados foram usados. Exigirá, também, treinamento adequado aos operadores de software que inserem os dados, manipulam ou interpretam os resultados, e esse treinamento deve estar institucionalmente incorporado nas rotinas do controlador. O treinamento deve dedicar atenção específica às limitações dos algoritmos, especialmente à possibilidade de falsos positivos e vieses de automação, bem como às responsabilidades individuais e institucionais na interpretação dos resultados.

De fato, o problema dos vieses é um dos mais preocupantes, uma vez que as visões e crenças dos desenvolvedores (ainda predominantemente homens brancos) podem transparecer no algoritmo. Conjuntos de dados também podem ser tendenciosos e direcionar decisões desproporcionalmente a grupos minoritários (viés de *big data*), como aconteceu com o Compass nos EUA, em que pessoas negras e oriundas de bairros pobres que concentram minorias eram consideradas mais propensas a reincidir e por isso tinham negado o direito de responder ao processo penal em liberdade. Daí que a qualidade (senão a própria justiça) das decisões dependerá de um controle rígido e adequação dos dados imputados para o treinamento e a operabilidade dos sistemas.

Finalmente, é fundamental reconhecer que os resultados produzidos pela inteligência artificial não passam de estimativas probabilísticas e, para prevenir injustiças, é essencial que a supervisão humana prevaleça, assegurando que decisões impactantes na vida das pessoas

sejam sempre ratificadas por seres humanos. As inferências algorítmicas devem ser vistas como hipóteses sujeitas a ajustes, e não como verdades incontestáveis. Diante disso, é imperativo manter um sistema de vigilância que acompanhe o desenvolvimento e a aplicação dessas tecnologias, alinhado aos valores éticos da IA, garantindo que qualquer restrição de direitos oriunda de decisões automatizadas possa ser revisada por indivíduos, conservando o humano no controle das escolhas tecnológicas.

Ao navegar pela nova era da inteligência artificial, transparência, responsabilidade e qualidade dos dados emergem como pilares essenciais para a mitigação dos riscos, não apenas como princípios retóricos, mas como verdadeiros alicerces para a construção de um ecossistema digital ético. A sinergia entre uma governança robusta e a vigilância contínua assegura que a IA se desenvolva de forma responsável, mantendo o respeito pelos direitos humanos e pela dignidade. A implementação de práticas positivas de uso dos sistemas de IA, ancoradas na responsabilização e em programas de governança consistentes é fundamental. Só assim teremos um progresso seguro e equitativo, pavimentando o caminho para um futuro em que a tecnologia avança em harmonia com os valores sociais e individuais.

Fonte: https://conjur.jumps.com.br/2024-jan-02/governanca-em-inteligencia-artificial-riscos-e-fatores-minimos-de-prevencao/