

Núcleo de Referência em Inteligência Artificial Ética e Confiável

Contribuição à consulta pública sobre o marco regulatório da Inteligência Artificial

O Núcleo de Referência em Inteligência Artificial Ética e Confiável aqui apresenta suas observações no contexto da Consulta Pública sobre o marco regulatório da Inteligência Artificial (IA). A presente contribuição aborda o eixo temático 2 - Impactos da IA, sobretudo o item 2.2.2 - Potencial Discriminatório, também apresentando sugestões pertinentes aos temas 3 e 4 (Direitos e deveres, e Accountability, Governança e Fiscalização).

Identificamos o conceito da **categorização e hierarquização dos riscos** (contemplando, por exemplo, os chamados “riscos inaceitáveis” da IA¹) como chave para a regulação sobre IA, traduzindo preocupações gerais inerentes. Evidentemente, a referência respectiva no marco regulatório há de ser principiológica e indicar categorias básicas, a serem futuramente monitoradas e detalhadas, por ocasião da regulamentação da lei, assim como tem ocorrido com outras normas, como a lei do Marco Civil da Internet, e mais recentemente, com a LGPD. Essa premissa serve ao propósito fundamental de distinguir - e tratar proporcionalmente - os diversos graus de impacto potencial da IA para os direitos humanos.

Neste aspecto, a lei brasileira pode se valer de boas práticas geralmente aceitas, como a proposta europeia de normatização², que se baseia na gradação de risco para estabelecer obrigações técnicas e legais crescentes, e a taxonomia preconizada pela OCDE³ para classificação e avaliação de sistemas de IA, que fomenta a universalização de critérios para políticas de IA. Notar tais referências há de ensejar uma plataforma legal mais amadurecida, e favorecer a inserção harmônica de sistemas de IA brasileiros no panorama global.

A filosofia do grau de risco repercute nos projetos de lei sobre IA ora considerados, por exemplo no PL 5051/2019 Art 4º - § 2º, que prevê responsabilização do seu supervisor, com perfil indefinido e sem distinguir situações básicas em que sua nomeação seja efetivamente viável ou justificável⁴.

Por igual, a IA afeta pautas reguladas por outras leis, como o Código do Consumidor, a LGPD e outras. Assim, uma indicação razoável deve ser incluída, como determinação para que a posterior regulamentação e atuação das autoridades competentes cuidem desta implicação.

¹ Fonte: Agência Senado

<https://www12.senado.leg.br/noticias/materias/2022/04/04/sugestoes-ao-marco-regulatorio-da-ia-sera-o-recebidas-ate-13-de-maio>

²<https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence>

³ <https://oecd.ai/en/classification>

⁴ Sistemas que sempre requeiram supervisão humana não devem ser considerados como de IA. Imagine-se, por exemplo, o contrassenso da hipótese de sistema dito de IA para filtrar mala direta eletrônica (spam), que sempre requeira intervenção humana para a análise do conteúdo.

Formulamos a seguir algumas propostas de contribuição, divididas por tópicos, e ao final, agrupamos algumas premissas a título de conclusão.

Propostas (para a lei de IA e para a sua regulamentação):

i) O tema da discriminação algorítmica em IA, em especial em sua sub-área - Aprendizado de Máquina, deve ser regulado levando em conta os papéis desempenhados pelos desenvolvedores e produtores da tecnologia e os direitos dos respectivos clientes e usuários. Outras leis já seguem essa prática, como a LGPD, que protege os titulares de dados, e distingue os papéis de Controladores, Operadores e Encarregados, e o Código do Consumidor, que enfoca consumidores (inclusive pessoas expostas a ofertas e riscos) e diversos tipos de fornecedores (produtores, comerciantes, etc);

ii) Há modalidades de discriminação *ex-ante*, ou seja, inspiradas por viés discriminatório antes da materialização em algoritmo; por isto, desenvolvedores e usuários de tecnologias digitais devem ser previamente cientificados de considerações éticas, padrões de boas práticas e responsabilidades respectivas, por meio de políticas, treinamento educativo ou certificação, procurando evitar discriminações preconcebidas; termos de ajustamento de conduta e responsabilização administrativa ou legal devem ser associados aos casos de maior risco e gravidade, juntamente com a exigência de transparência sobre bases de dados acessadas no desenvolvimento ou uso da IA;

iii) Há também modalidades de discriminação *ex-post*, ou seja, incorporadas no desenvolvimento ou utilização de viés algorítmico; o tratamento deve ser similar ao adotado quanto às discriminações preconcebidas (descrito no item ii acima), adicionando-se para os casos de risco mais acentuado a exigência da figura do Encarregado de IA, como responsável pela coordenação e atendimento de pleitos de autoridades ou de terceiros interessados (semelhantemente à figura do Encarregado de dados, presente na LGPD), a previsão de relatório de impacto de IA, e a estipulação de multa administrativa com faixa de valores com capacidade de dissuasão de práticas nocivas; a preservação de segredos comerciais ou industriais não deve ser impeditiva da transparência nos casos em que esta for imperativa;

iv) Um glossário deve constar na lei, para educação da população sobre termos técnicos fundamentais e para equalização informativa capaz de subsidiar interpretação e aplicação;

v) A classificação de riscos deve observar categorias conceituais básicas, como no modelo europeu, concentrado em três categorias essenciais de risco: inaceitável, elevado e mínimo;

vi) A classificação de riscos deve ser pautada pela consideração dos seguintes fatores: explicitação do objetivo e escopo do sistema de IA; proteção da segurança do sistema de IA; grau de implicação quantitativa e qualitativa para usuários e população em geral (inclusive disponibilidade de relatório de impacto de IA e informação sobre limitações de previsibilidade de operações e resultados do sistema de IA); programa educativo fomentador de cultura de responsabilidade por IA; canal de esclarecimentos de dúvidas e atendimento de solicitações; adoção de práticas de ética "*by design*"/nativas no

desenvolvimento de sistemas de IA; comunicação proativa de ameaças de riscos graves; realização de testes prévios; equipe multidisciplinar (e representativa de diversidade, quando inerente à natureza do sistema de IA); disponibilidade de documentação técnica e de interfaces (APIs); sistemas de monitoramento e detecção de padrões discriminatórios ilícitos ou abusivos.

Considerações Finais

A regulação de uma tecnologia em contínuo desenvolvimento e que pode oferecer riscos, como a IA, é desejada e necessária. No entanto, um marco normativo sobre algo que ainda não conseguimos definir com maior propriedade é algo complexo. O mundo todo tem buscado soluções para esta moldura legal. O estudo mais amadurecido que se conhece é o da União Europeia, consolidado na versão divulgada em abril de 2021², e que é resultado de longo processo de discussão e decantação.

O comedimento na regulação preliminar do assunto no Brasil, procurando uma justa medida com objetivo de proteger a população usuária contra os riscos mais acentuados, porém sem inibir a oferta de benefícios relevantes, justifica-se inclusive pelo fato de que o país já conta com leis correlatas (como a LGPD e o Código do Consumidor) passíveis de integração na aplicação da disciplina legal do assunto, a exemplo de no tópico aqui já abordado, de cuidados contra tratamentos discriminatórios ilícitos ou abusivos.

Uma regulação com conteúdo precipitado pode ser causa de atraso mais do que de avanço. A tecnologia nacional de IA merece um porto seguro para investimento e desenvolvimento, em sintonia com os anseios sociais e os parâmetros legais.

O Núcleo de Referência em Inteligência Artificial Ética e Confiável, grupo formado por cientistas e profissionais devotados ao assunto, sem fins lucrativos, se coloca à disposição dos legisladores e das comissões de assessoramento para subsidiar os debates e a atividade legiferante, nesta etapa e nas que hão de seguir-se.

Membros do Núcleo de Referência em Inteligência Artificial Ética e Confiável autores deste estudo:

Ana Cristina Bicharra Garcia (Universidade Federal do Estado do Rio de Janeiro - UNIRIO)

Elizabeth Maria Freire de Jesus (Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais/UFRJ)

Gilberto Martins de Almeida (Instituto de Direito e Tecnologia - IDTEC)

Harold Dias de Mello Junior (Universidade do Estado do Rio de Janeiro - UERJ)

João Carlos Pereira da Silva (Universidade Federal do Rio de Janeiro - UFRJ)

John Lemos Forman (J.Forman Consultoria)

Luciana Cardoso de Castro Salgado (Universidade Federal Fluminense - UFF)

Mariza Ferro (Laboratório Nacional de Computação Científica - LNCC)

Márjory Cristiany Da Costa-Abreu (Sheffield Hallam University, Reino Unido)

Marlo Vieira dos Santos e Souza (Universidade Federal da Bahia - UFBA)

Marcio Cordeiro Galvão (Módulo Security Solutions S.A.)

Priscila Machado Vieira Lima (Universidade Federal do Rio de Janeiro - UFRJ)

Roberto Pinto Souto (Laboratório Nacional de Computação Científica - LNCC)