

Instituto de Tecnologia
e Sociedade
Abril, 2024

Como regular a

Inteligência Artificial?

Expandindo os horizontes de análise
para além da União Europeia



Instituto
de Tecnologia
& Sociedade
do Rio



Resumo Executivo

O presente relatório oferece uma introdução aos diferentes modelos de regulação de Inteligência Artificial (IA) de forma a facilitar a compreensão das nuances envolvidas nas discussões sobre o tema no Brasil, buscando expandir os horizontes de análise para além da União Europeia. São identificados, ao todo, três macro-modelos de regulação.

- Primeiro, o **modelo de incorporação**, baseado em normas já existentes que disciplinam, direta ou indiretamente, aplicações específicas de IA.
- Segundo, o **modelo emergente**, baseado na criação de novas normas para disciplinar de forma específica a IA.
- Terceiro, o **modelo ético**, baseado na articulação de princípios éticos para informar o desenvolvimento e aplicação da IA.

Esses modelos, ao seu turno, dependem do funcionamento de diferentes ferramentas regulatórias que podem ser combinadas de diferentes formas. Sem buscar esgotar a discussão, o relatório foca em algumas ferramentas de maior proeminência: investigação de algoritmos, auditoria externa, análise de risco, sandboxes regulatórios, requerimentos de transparência, direitos dos usuários, cooperação multissetorial e código de conduta.

Ao final, são apresentadas algumas iniciativas regulatórias para além do Ato da IA da União Europeia, em especial no Japão, nos Estados Unidos e no Reino Unido, passando, também, por algumas iniciativas transnacionais e internacionais. O relatório chega à conclusão de que é necessário **resistir à tentação de assumir um ou outro modelo como paradigmático** antes mesmo que se tenha a chance de testar diferentes soluções e alternativas em contextos distintos.

Introdução: Um novo panorama regulatório

Os últimos anos foram marcados por diferentes iniciativas regulatórias envolvendo o funcionamento e os impactos das novas tecnologias na sociedade. Foi o caso, por exemplo, das legislações sobre proteção de dados e privacidade na segunda metade dos anos 2010 – como a *General Data Protection Regulation* na União Europeia e a Lei Geral de Proteção de Dados no Brasil – e das regulações de plataformas digitais a partir de 2020 – como o *Digital Services Act* na União Europeia e o PL nº 2.630 no Brasil. Vale destacar que essas e outras iniciativas regulatórias já contam com dispositivos que se relacionam, direta ou indiretamente, com o desenvolvimento e implementação de aplicações de Inteligência Artificial (IA). Nada obstante, hoje estamos diante de uma nova onda regulatória que busca disciplinar a IA de forma específica. Veja-se, nesse sentido, o *Artificial Intelligence Act* em fase final de negociações na UE e o PL nº 2.338/2023 no Brasil.

Em geral, as discussões sobre a IA são pautadas pelos riscos representados pela tecnologia. Exemplos não faltam, como o uso de aplicações baseadas em IA para a confecção de vídeos e imagens hiper-realistas que podem ter um impacto incomensurável no processo eleitoral¹ ou sobre a honra e a imagem das pessoas². Essa também emerge como a nova fronteira das fraudes financeiras na Internet – em fevereiro de 2024, por exemplo, um funcionário de uma multinacional foi enganado por criminosos que usaram um *deepfake* para se passar pelo diretor financeiro da companhia em uma videoconferência, resultando numa transferência de 25 milhões de dólares aos golpistas³. Esses casos que recebem maior atenção midiática, entretanto, são apenas a ponta do iceberg. A IA, afinal, não se resume ao campo da IA Generativa e está presente em atividades do cotidiano como análise e aprovação de crédito, atendimento ao cliente, recomendação de produtos e serviços em plataformas digitais e avaliação de currículos.

Como ensina Kate Crawford, há uma significativa confusão conceitual em torno da IA, em especial pelo distanciamento gerado pela falsa percepção de que se trata de um campo completamente técnico, inacessível para pessoas não treinadas em computação e linguagens de programação⁴. Pelo contrário, como afirma a autora, “a IA é fundamentalmente política”⁵, no sentido de que seu desenvolvimento depende de recursos naturais e humanos, não sendo possível desvincular a tecnologia do seu contexto sociopolítico. A proposição de Crawford, então, é de que a IA seja compreendida como uma espécie de atlas, ou seja, uma representação contingente do mundo que é moldada, inevitavelmente, por forças sociais, políticas e econômicas⁶. Ainda, segundo a visão de Yuk Hui, é preciso quebrar com a perspectiva de que a IA é “universal” e, conseqüentemente, de que seu destino final seria a formulação de

1 MARTINS, Américo. Eleições nos EUA: uso de deepfake e IA revela problema que pode se repetir no Brasil. **CNN Brasil**, 23 de janeiro de 2024. Acesso em 5 de fevereiro de 2024.

2 PANCINI, Laura. O que são deepnudes? Taylor Swift é vítima de imagens pornográficas geradas por IA. **Exame**, 26 de janeiro de 2024. Acesso em 5 de fevereiro de 2024.

3 CHEN, Heather. MAGRAMO, Kethleen. Funcionário de multinacional paga US\$ 25 mi a golpista que usou “deepfake” para simular reunião. **CNN Brasil**, 4 de fevereiro de 2024. Acesso em 5 de fevereiro de 2024.

4 CRAWFORD, Kate. **Atlas of AI: power, politics, and the planetary costs of artificial intelligence**. New Haven: Yale University Press, 2021.

5 *Ibid.*, p. 9.

6 *Ibid.*, p. 9-14.

uma superinteligência artificial apta a substituir o ser humano em todas as suas funções e atribuições⁷. Afinal, inexiste uma concepção universal de inteligência⁸ e a tecnologia, ao seu turno, deve sempre ser pensada a partir de uma pluralidade de cosmovisões⁹.

Essas considerações trazem uma nova e bem-vinda complexidade ao debate regulatório. Não há de antemão um modelo correto ou superior de regulação da IA. Existem, na prática, diferentes modelos de regulação que se valem da combinação de distintas ferramentas, como análises de risco, sandboxes regulatórios e requerimentos de transparência. É preciso resistir à tentação de se assumir um ou outro modelo como paradigmático antes mesmo que se tenha a chance de testar diferentes soluções e alternativas em contextos distintos, as quais, vale ressaltar, muitas vezes necessitam de abordagens minimamente personalizadas para que sejam bem-sucedidas. Assim, a forma como uma determinada jurisdição opta por regular essa nova tecnologia importa sobremaneira e esse será justamente o objeto de análise do presente relatório. Trata-se, portanto, de uma primeira aproximação com esse debate ainda em pleno desenvolvimento. Antes de avançar, entretanto, dois pressupostos precisam ser considerados e esclarecidos: o que é inteligência artificial e qual é o objetivo que se busca alcançar com a sua regulação pelo estado.

Em linhas gerais, para os fins deste relatório, “a IA é um sistema baseado em máquinas que pode, com base em um conjunto de objetivos definidos por humanos, fazer previsões, recomendações ou decisões que influenciam ambientes virtuais ou físicos”¹⁰. Enquanto aplicação tecnológica, a IA está inserida no nosso cotidiano há décadas, além de estar presente em debates acadêmicos desde pelo menos o início da década de 50 do século passado¹¹. Entretanto, é inegável a aceleração em seu desenvolvimento nos últimos anos e o lugar que essa tecnologia agora ocupa no imaginário social. Independentemente da forma como a IA será regulada na prática, é preciso ter em mente que o

7 HUI, Yuk. *On the Limit of Artificial Intelligence*. **Philosophy Today**, v. 65, n. 2, 2021.

8 *Ibid.*, pp. 351-54.

9 HUI, Yuk. **Tecnodiversidade**. São Paulo: Ubu Editora, 2020, pp. 89-96.

10 Essa é a definição adotada pela OCDE. Para maiores detalhes, ver <<https://mneguidelines.oecd.org/RBC-and-artificial-intelligence.pdf>>.

11 ARHEGAS, João Victor. MAIA, Gabriella. O que é inteligência artificial (IA)? Análise em três atos de um conceito em desenvolvimento. **Cadernos Adenauer**, v. XXIII, n. 2, 2022, pp. 14-17.

objetivo da regulação é (e sempre deve ser) “compatibilizar o desenvolvimento e a implementação da IA com valores e princípios humanos” de forma que “a IA [beneficie e não ameace] a humanidade em suas diversas dimensões, desde relações de trabalho até o exercício da democracia”¹².

¹² *Ibid.*, p. 11.

1. Diferentes Modelos de Regulação da Inteligência Artificial

A regulação em discussão atualmente no Congresso Nacional, o PL nº 2.338 de 2023, é resultado direto de um anteprojeto elaborado por uma comissão de juristas encabeçada por referências na área como Ricardo Villas Bôas Cueva, ministro do Superior Tribunal de Justiça, e Laura Schertel Mendes, professora do IDP e da UnB. O trabalho da comissão foi fortemente influenciado pelo Ato da IA (ou AI Act) que está em fase final de negociações na União Europeia e deve se tornar a primeira regulação específica dessa natureza no mundo. A regulação europeia, por sua vez, orbita em torno de um mecanismo de análise de risco, focando não no desenvolvimento da tecnologia de um ponto de vista técnico, mas sim nas suas potenciais externalidades negativas. Os riscos cobertos pelo Ato da IA são aqueles que se relacionam com as áreas da saúde e da segurança, além daqueles que podem impactar negativamente o exercício de direitos fundamentais no território europeu.

O documento menciona quatro níveis de risco que, por sua vez, estão associados a diferentes obrigações regulatórias¹³. Em primeiro lugar, aplicações de risco inaceitável, como, por exemplo, a análise de crédito social, são proibidas tendo em vista seus impactos desproporcionalmente negativos para os cidadãos. Em segundo lugar, aplicações de alto risco exigem uma autorização

¹³ Ver, nesse sentido, o Anexo III do Ato da IA da União Europeia para uma lista completa dos sistemas classificados como de alto risco: <<https://www.euaiact.com/annex/3>>.

prévia para sua implementação. A grande maioria dos casos listados pela regulação se inserem nessa segunda categoria, como, por exemplo, a análise de crédito financeiro. Em terceiro lugar, aplicações de risco limitado são vinculadas a exigências de transparência, em especial a respeito de como é feito seu treinamento. Se inserem aqui, por exemplo, os chatbots. Por fim, em quarto lugar, aplicações de baixo risco não sofrem restrições específicas, como, por exemplo, jogos de computador que se valem de funções desempenhadas por IA, como renderização de imagens e reprodução de linguagem natural. A iniciativa da UE, vale ressaltar, não se reduz à análise de risco, mas esse é, inegavelmente, o elemento mais influente do Ato da IA em outras jurisdições que também se debruçam sobre o tema.

O PL nº 2.338, ao seu turno, trabalha com dois níveis de risco. Antes da sua colocação no mercado, todo sistema de IA deverá passar por uma avaliação preliminar pelo fornecedor da tecnologia para que seu grau de risco seja classificado. Algumas aplicações, nos termos do projeto, seriam vedadas dado o risco inaceitável envolvido em sua implementação. As hipóteses de proibição estão previstas em seu Art. 14, incluindo, por exemplo, sistemas de IA “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” (Art 14, inciso I). Outras aplicações são classificadas como de alto risco, sendo obrigatória a realização de uma “avaliação de impacto algorítmico”. As hipóteses de IA de alto risco estão listadas em seu Art. 17 e incluem gestão de infraestrutura crítica, recrutamento de candidatos para uma vaga de emprego, análise de crédito financeiro, funcionamento de veículos autônomos, administração da justiça, condução de investigação criminal, dentre outras.

Em que pese a utilidade e eficiência dessa moldura, a análise de risco é apenas uma dentre tantas outras ferramentas disponíveis ao regulador. Mais do que isso, o modelo que propõe a criação de uma regulação específica de IA também é um dentre outros possíveis, sendo necessário colocar em primeiro plano suas vantagens e desvantagens para que se possa compreender as consequências de sua escolha em detrimento dos demais. Veja-se, assim, que falar apenas em regulações específicas de IA que orbitam em torno de classificações de risco é confundir a árvore com a floresta. Isso se deve, em parte, ao que Anu Bradford chamou de “Efeito Bruxelas”, destacando a influência

sem precedentes das políticas públicas europeias em outros países, especialmente aquelas que envolvem a regulação de novas tecnologias¹⁴. A discussão que segue, assim, é uma primeira tentativa de indicar possíveis alternativas que não podem ser desconsideradas de bate-pronto, trazendo uma maior nuance ao debate regulatório nacional.

Antes disso, entretanto, é preciso deixar claro que os modelos apresentados a seguir não devem ser encarados como auto-excludentes. É dizer, a opção por um modelo não exclui a possibilidade de outro também ser adotado de forma complementar em uma determinada jurisdição. Assim, sugere-se que esses modelos sejam encarados como lentes que podem ser sobrepostas de forma a individualizar a abordagem regulatória da IA. Diferentes países combinam diferentes modelos em um único esquema regulatório a partir da implementação de múltiplas ferramentas (as quais serão analisadas no capítulo seguinte), dando maior peso para alguns em detrimento de outros. Ainda, é preciso ressaltar que os modelos aqui apresentados são amplos e abrangentes e foram identificados a partir da experiência regulatória que envolve não apenas a disciplina da IA ao redor do mundo como também de novas tecnologias em geral. São, portanto, modelos ideais que podem sofrer mudanças específicas a depender do contexto no qual são implementados e para quais finalidades. Ao todo são identificados três macro-modelos: o modelo de incorporação, o modelo emergente e o modelo ético¹⁵.

2.1. Modelo de Incorporação

| O QUE É? | VANTAGEM | DESvantagem |
|---|--|---|
| Foca em normas já existentes e como elas se relacionam com diferentes aplicações de IA. | Evita redundância ou sobreposição regulatória, garantindo maior segurança jurídica e não implicando em custos regulatórios desnecessários. | A descentralização pode gerar inconsistências e, se não atualizado de forma setorial, pode se tornar irrelevante rapidamente. |

O modelo de incorporação – que também pode ser referido, em linguagem jurídica, como modelo de subsunção – orbita em torno de normas, legislações e regulações já existentes no ordenamento jurídico e suas possíveis interseções

¹⁴ BRADFORD, Anu. **The Brussels Effect: How the European Union Rules the World**. Oxford: Oxford University Press, 2020.

¹⁵ Para uma análise acadêmica de diferentes formas de regulação da IA por meio do Direito, ver PETIT, Nicolas. COOMAN, Jerome De. *Models of Law and Regulation for AI*. **EUI Working Papers**, RSCAS n. 63, 2020.

com casos que envolvam o uso de IA na sociedade. Nas palavras de Nicolas Petit e Jerome De Cooman, o modelo em questão propõe uma abordagem baseada na “identificação do direito relevante, contra o qual são confrontadas questões de fato envolvendo um sistema de IA. Na prática, a questão de fato geralmente é um caso de uso de IA que leva à uma disputa”¹⁶. Veja-se, a título de exemplo, as diversas questões levantadas por modelos de IA Generativa como o ChatGPT no campo do Direito Autoral, desde o uso de obras protegidas para o treinamento do modelo até a confecção de novas obras a partir dos inputs recebidos.

Esse modelo, portanto, explora o potencial do direito já existente para disciplinar o uso de IA em setores específicos¹⁷. Não é incomum, inclusive, que alguns desses setores já disponham de normas específicas sobre aplicações de IA que lhes dizem respeito, como é o caso do uso de IA para análise de crédito no sistema financeiro e bancário. Nesses casos, a criação de uma regulação específica de IA como o PL nº 2.338/2023 pode gerar uma indesejável sobreposição normativa que, por sua vez, será responsável por um aumento da insegurança jurídica e dos custos associados à regulação. Assim, o principal ponto positivo do modelo de incorporação é trazer maior eficiência normativa ao ordenamento jurídico, evitando redundância ou sobreposição regulatória, além de colocar no radar da agenda regulatória as especificidades de cada setor. O ponto negativo, entretanto, é que uma maior descentralização regulatória pode gerar inconsistências com o tempo e, se não for atualizada à luz dos novos desdobramentos tecnológicos, corre o risco de se tornar irrelevante rapidamente. Elementos desse modelo, como se verá abaixo, estão presentes nas abordagens regulatórias de países como Japão e Estados Unidos.

¹⁶ *Ibid.*, p. 2. Tradução livre.

¹⁷ Para uma análise das bases normativas já existentes no ordenamento jurídico sobre IA, ver o relatório “Panorama Regulatório de Inteligência Artificial no Brasil” do ITS Rio. Disponível em <<https://itsrio.org/pt/publicacoes/panorama-regulatorio-de-inteligencia-artificial-no-brasil/>>.

2.2. Modelo Emergente

| O QUE É? | VANTAGEM | DESVANTAGEM |
|---|--|---|
| Foca na criação de uma nova regulação, a partir de uma legislação específica, com o intuito de disciplinar os efeitos emergentes da IA. | Confere maior atualidade e relevância à regulação, incentivando que o regulador enfrente questões relacionadas à tecnologia de forma específica. | Pode promover redundância ou sobreposição regulatória, além de desconsiderar especificidades setoriais. |

O modelo emergente sugere, por sua vez, a criação de um novo campo do Direito que possa se debruçar de forma específica sobre o desenvolvimento e aplicação de sistemas de IA na sociedade. Como explicam Petit e De Cooman, esse modelo parte do pressuposto de que “sistemas de IA produzem fenômenos emergentes” com efeitos únicos para a sociedade e, por isso, devem ser disciplinados a partir da criação de um novo braço do direito¹⁸. Um marco desse modelo é o foco na tecnologia em si já que, em geral, o direito existente é prontamente desconsiderado ou, ao menos, colocado em segundo plano. É dizer, o foco passa a ser o que há de específico no funcionamento da IA e suas consequências jurídicas, como a criação de direitos para os sujeitos que interagem com sistemas de IA em diferentes circunstâncias do cotidiano. É o caso, por exemplo, do direito de ser informado de que o atendimento virtual referente a uma compra online está sendo feito por um chatbot e não por um atendente humano, principalmente quando o uso de linguagem natural é sofisticado e baseado em modelos de IA Generativa.

O principal ponto positivo do modelo emergente é sua atualidade e relevância, permitindo ao regulador que considere o estado da arte da tecnologia e suas implicações jurídicas. Ademais, trata-se de um modelo que conduz ao enfrentamento de questões importantes que muitas vezes podem ser descartadas como “utópicas” ou “apressadas” à luz de outros modelos, a exemplo da discussão sobre a possibilidade de se atribuir personalidade jurídica aos sistemas de IA. Entretanto, o ponto negativo é que o modelo emergente pode causar redundância ou sobreposição regulatória, aumentando o custo de adequação para determinados atores e aumentando sua insegurança jurídica. Ademais, é um modelo que tende a desenvolver uma espécie de “visão

¹⁸ *Ibid.*, p. 4. Tradução livre.

de túnel”, focando apenas em sistemas de IA de uma forma geral e perdendo de vista as especificidades envolvidas na implementação da tecnologia em diferentes contextos e setores. Esse é, em linhas gerais, o modelo que prevalece atualmente na União Europeia e no Brasil a partir, respectivamente, do Ato da IA e do PL nº 2.338 de 2023.

2.3. Modelo Ético ou Principiológico

| O QUE É? | VANTAGEM | DESvantagem |
|---|--|---|
| Foca na estruturação de princípios para guiar o desenvolvimento e a implementação da IA, evitando normas específicas que podem tornar a moldura desatualizada em pouco tempo. | Maior flexibilidade em razão dos princípios de textura aberta, mantendo a moldura regulatória atualizada e funcional por mais tempo. | Dificuldade de se atingir um consenso mínimo sobre valores éticos para guiar a regulação da tecnologia, além de uma potencial falta de aderência diante do enfraquecimento de mecanismos coercitivos. |

Por fim, em terceiro e último lugar, o modelo ético ou principiológico é uma forma de regulação mais branda da IA baseada na criação e promoção de princípios para o desenvolvimento e implementação da tecnologia, evitando, assim, normas específicas. Essa é uma discussão presente no campo de regulação de novas tecnologias há décadas, como no caso dos debates sobre governança da Internet a partir dos anos 90. É dizer, devemos criar regras específicas e rígidas sobre governança da Internet ou apostar no desenvolvimento de princípios flexíveis e de textura aberta? Como esclarecem Petit e De Cooman, “sistemas de IA tocam principalmente em um campo da ética conhecido como ética normativa” e o seu propósito é “criar normas morais para distinguir o que é bom do que é ruim” em termos de desenvolvimento e implementação da tecnologia¹⁹. Nesse sentido, podemos pensar em princípios como a participação humana em todas as fases de vida de um sistema de IA ou então a necessidade de se identificar o ator responsável por sua implementação, manutenção ou distribuição.

Um exemplo dessa abordagem é o estudo de Frank Pasquale no qual o autor propõe a atualização das “Leis da Robótica” para a era da IA²⁰. Pasquale

¹⁹ *Ibid.*, p. 5. Tradução livre.

²⁰ PASQUALE, Frank. **New Laws of Robotics: Defending Human Expertise in the Age of AI**. Cambridge: Belknap Press, 2020.

indica alguns princípios éticos que devem guiar nossas discussões, argumentando, por exemplo, que sistemas de IA não devem “imitar ou falsificar” a humanidade e que a tecnologia deve ser desenvolvida para, em um primeiro momento, complementar ao invés de substituir trabalhadores. Esse tipo de iniciativa informada por um modelo ético de regulação de IA tem como ponto positivo uma maior flexibilidade normativa, o que é bem-vindo e muitas vezes necessário quando o objeto a ser regulado é uma tecnologia emergente em constante evolução e transformação. Do outro lado, é possível destacar como um ponto negativo desse modelo a dificuldade de se atingir um consenso mínimo sobre o conteúdo de seus comandos éticos e valorativos – muito em razão da sua alta carga de subjetividade – e uma possível falta de aderência diante da textura aberta dos princípios e do consequente enfraquecimento de eventuais mecanismos de coerção. Elementos desse modelo, como se verá abaixo, podem ser identificados em países como o Reino Unido.

2. Caixa de Ferramentas para Regulação da Inteligência Artificial

Como mencionado anteriormente, os três modelos acima são amplos e abrangentes, configurando, assim, modelos ideais que podem sofrer modificações a depender do contexto nos quais são implementados e para quais finalidades. Essas mudanças podem acontecer de duas principais formas. Primeiro, como não são auto-excludentes, é possível que elementos de diferentes modelos sejam sobrepostos, funcionando como lentes dentro de um modelo próprio de regulação. Segundo, a maior parte da personalização dentro dos modelos se dá pela escolha entre diferentes ferramentas que podem ser combinadas para operacionalizá-lo. O presente capítulo, assim, tem por objetivo apresentar algumas das principais ferramentas de regulação da IA que podem ser implementadas pelos diferentes modelos a fim de se atingir determinadas finalidades regulatórias.

3.1. Investigação de Algoritmos e Auditoria Externa

Um dos principais pontos discutidos na regulação de IA – e de novas tecnologias em geral, como também o caso da governança de plataformas digitais de redes sociais – é a falta de transparência em sistemas algorítmicos. Em uma das principais obras sobre o tema, Frank Pasquale fala em uma “sociedade da caixa preta” para se referir à opacidade no funcionamento de algoritmos que “decidem” questões importantes para nossa vida em sociedade, desde qual conteúdo será recomendado em uma rede social até qual será o limite máximo do cartão de crédito no aplicativo do banco²¹. Assim, uma ferramenta comum em iniciativas regulatórias é a promoção de investigações e auditorias em sistemas de IA.

Vale destacar, desde já, que essa ferramenta é alvo de críticas e questionamentos, em especial por, a depender de como for implementada, colocar em risco o segredo comercial que protege sistemas algorítmicos no mercado de novas tecnologias. Dito isso, é possível dividir essa ferramenta em duas frentes. Em primeiro lugar temos a investigação (ou inspeção) técnica de algoritmos, que se debruça sobre os inputs e outputs de um determinado sistema buscando testar uma hipótese – por exemplo, “avaliando se o sistema exibe sinais de enviesamento racial” em sua interação com o usuário²². Em segundo lugar temos a auditoria externa exigida em regulação, que serve para “identificar as consequências sociais do uso de um sistema e avaliar o seu funcionamento à luz de um parâmetro normativo previamente estabelecido, de forma a mapear potenciais áreas de preocupação”²³.

3.2. Análise de Risco e Proibição

A análise de risco é a ferramenta que foi adotada para servir de centro gravitacional em regulações como o Ato da IA na União Europeia e o PL nº 2.338/2023 no Brasil. Seu objetivo é duplo: identificar riscos no funcionamento de sistemas

²¹ PASQUALE, Frank. **The Black Box Society**: *The Secret Algorithms That Control Money and Information*. Cambridge: Harvard University Press, 2016.

²² *Ibid.*, p. 8. Tradução livre.

²³ *Ibid.*, p. 8. Tradução livre.

de IA e, a partir disso, criar obrigações regulatórias para mitigá-los sempre que possível – além de, muitas vezes, estar associada à possibilidade de responsabilização legal em caso de negligência seguida da concretização dos riscos previamente identificados²⁴. Veja-se que essa é uma ferramenta que comumente estará associada a outras, como a imposição de auditorias externas para que seja possível determinar se eventuais medidas de mitigação estão sendo tomadas da forma adequada ou não.

Em alguns casos, entretanto, o risco identificado pode ser alto demais a ponto de se justificar a proibição de um determinado sistema da IA; por exemplo, “proibições e moratórias foram aplicadas às tecnologias de reconhecimento facial por departamentos de polícia e até mesmo governos locais nos EUA, incluindo Portland, Oakland e São Francisco”²⁵. Essas ferramentas, vale dizer, têm a vantagem de concentrarem seus esforços sobre a aplicação da tecnologia ao invés do desenvolvimento da tecnologia em si, evitando, assim, maiores interferências no funcionamento do mercado e na promoção de inovação tecnológica no setor. Nada obstante, há também a desvantagem de não ser possível, de antemão, prever todos os riscos e tendências envolvidas no desenvolvimento da tecnologia, perdendo de foco, também, a tarefa igualmente importante de se potencializar aplicações positivas de IA e não apenas limitar ou proibir as negativas.

3.3. Sandboxes Regulatórios de Inteligência Artificial

Uma ferramenta promissora que vem sendo explorada por diferentes iniciativas regulatórias são os sandboxes de IA. Trata-se, em linhas gerais, da estruturação de um ambiente regulatório experimental sob supervisão de uma autoridade pública. É, assim, uma promissora técnica regulatória que permite à autoridade competente a testagem de novas abordagens regulatórias com um grupo pré-selecionado de desenvolvedores ou distribuidores de aplicações de IA. Tudo isso é feito em um ambiente controlado e com objetivos

²⁴ *Ibid.*, p. 7.

²⁵ *Ibid.*, p. 6. Tradução livre.

pré-definidos. Essa ferramenta está presente em diferentes propostas de regulação da IA, incluindo o Ato da IA e o PL nº 2.338/2023. Sua principal vantagem é a instalação de uma espécie de válvula de escape dentro do regime regulatório, a qual, por sua vez, permite melhor controlar a pressão dentro do sistema como um todo diante do surgimento de inovações tecnológicas não previstas inicialmente pelo modelo.

3.4. Relatórios de Transparência e Direitos dos Usuários

Outra ferramenta comumente utilizada para corrigir a opacidade no funcionamento de novas tecnologias, em especial aquelas que envolvem a implementação de algoritmos e técnicas de automatização, é a publicação de relatórios de transparência. Essa é uma técnica há muito debatida, por exemplo, no contexto de governança de plataformas digitais, buscando a publicação de informações-chave sobre a operação de mecanismos e protocolos de moderação de conteúdo em redes sociais. Em geral, entretanto, a exigência de publicação de relatórios de transparência não pode ser encarada como um fim em si mesma. Afinal, transparência não se traduz automaticamente em cognição, especialmente considerando que diferentes públicos terão diferentes graus de conhecimento sobre o funcionamento de sistemas de IA.

É preciso pensar, assim, em quais são as finalidades que podem ser atingidas a partir da publicação dos relatórios. São dois os principais motivos para isso. Em primeiro lugar, uma vez estabelecidas, essas finalidades informam a estruturação da exigência de transparência de forma específica; em outras palavras, quais dados serão requisitados e qual é a justificativa para a requisição? Em segundo lugar, é indispensável pensar nos direitos das pessoas afetadas por sistemas de IA e como a transparência ajudará a salvaguardá-los. Assim, por exemplo, é possível pensar, dentre outras, na seguinte finalidade: o empoderamento do usuário diante de decisões tomadas ou informadas por sistemas de IA que têm um impacto imediato sobre seus direitos fundamentais. Nesse cenário, a moldura regulatória poderá garantir ao usuário o direito de ser informado sempre que estiver interagindo com um sistema de IA, além da possibilidade de solicitar a revisão de decisões tomadas ou informadas por tais sistemas por um operador humano.

Veja-se, mais uma vez, que há uma certa circularidade e complementaridade entre as ferramentas mencionadas ao longo do presente capítulo. Ao definir os direitos das pessoas impactadas por sistemas de IA, o regulador terá uma melhor ideia de quais dados devem estar presentes nos relatórios de transparência. Em posse dessas informações, o regulador poderá verificar se tais direitos estão sendo respeitados e até mesmo, por meio de uma investigação mais aprofundada, identificar potenciais riscos envolvidos no funcionamento de determinados sistemas de IA. Isso, por sua vez, se conecta com as ferramentas de auditoria externa e análise de risco, mas também pode indicar que é o momento oportuno para se estruturar um sandbox e testar novas abordagens regulatórias com um determinado grupo de desenvolvedores dentro de um ambiente controlado.

3.5. Códigos de Conduta e Oversight

Por fim, é importante destacar a existência de ferramentas que priorizam uma abordagem de correção da IA em contraposição a uma regulação clássica pelo estado. Isso pode se dar, por exemplo, a partir da formulação de códigos de conduta com base em recomendações emitidas por um órgão regulador que, de preferência, será composto por representantes de diferentes setores da sociedade. O órgão poderá indicar, por exemplo, quais mecanismos e soluções de governança os desenvolvedores e fornecedores de IA podem adotar para mitigar determinados riscos envolvidos no uso da tecnologia. A partir do recebimento de tais recomendações, os desenvolvedores e fornecedores poderão desenvolver os seus próprios códigos de conduta com base nas especificidades das suas práticas internas e do seu posicionamento no mercado. O órgão regulador, ao seu turno, recebe os códigos depois de elaborados e poderá validá-los ou sugerir alterações. Isso força, na prática, um diálogo público entre estado e empresas na busca comum por soluções regulatórias efetivas, favorecendo, ainda, a estruturação de um sistema de supervisão ou oversight.

3. Diferentes Abordagens Regulatórias

Apresentados os diferentes modelos de regulação e a caixa de ferramentas à disposição dos reguladores, o presente capítulo será dedicado a uma breve discussão sobre diferentes iniciativas regulatórias de IA. Importa salientar, entretanto, que as iniciativas apresentadas a seguir não pretendem esgotar o tema, mas sim indicar algumas alternativas já em discussão que se distanciam, de uma forma ou outra, do Ato da IA da União Europeia.



4.1. Japão

Diferente da UE – que está prestes a adotar um modelo emergente de regulação da IA com foco em análise de risco –, o Japão vem adotando uma abordagem que se aproxima mais do modelo de incorporação mencionado acima. O governo japonês criou um Conselho Estratégico de Inteligência Artificial para discutir como o país deve se posicionar em debates regulatórios sobre o tema. O Conselho se reuniu pela primeira vez em maio de 2023 e é presidido pelo Professor Yutaka Matsuo, um especialista em IA da Universidade de Tóquio²⁶. Desde os primeiros encontros do Conselho já ficou evidente que as autorida-

²⁶ SHIMBUN, Yomiuri. *Japan's Artificial Intelligence Strategy Council Holds 1st Meeting with Aim to Create Rules for AI Use*. **The Japan News**, 11 de maio de 2023. Acesso em 7 de fevereiro de 2024.

des japonesas compartilham preocupações diferentes daquelas que dominaram os debates regulatórios na UE. Em uma entrevista para a Reuters em julho de 2023, Matsuo afirmou que as regras europeias são “restritivas demais” e que “na UE a questão é menos sobre como promover inovação e mais sobre como responsabilizar empresas já consolidadas”²⁷.

No final de maio de 2023, o Conselho publicou um documento chamado “Pontos Provisórios de Discussão sobre IA” no qual articula três principais argumentos que passaram a compor a base da estratégia regulatória do Japão. Em primeiro lugar, é preciso avaliar como leis e regulações já existentes se aplicam ao uso da IA, focando em diferentes setores como saúde, infraestrutura, educação, etc. O foco, portanto, deve ser em promover a aderência de desenvolvedores e fornecedores de IA aos parâmetros legais já em vigência no país. Em segundo lugar, apenas quando as normas existentes forem insuficientes é que se justifica a adoção de novas regulações sobre o tema, idealmente de forma flexível e setorial. Em terceiro e último lugar, o país não deve adotar como estratégia regulatória a criação de uma legislação específica para a IA como fez a UE, buscando, assim, uma maior flexibilidade regulatória à luz da inovação tecnológica.

Isso significa que, na prática, o Japão adotou como seu principal objetivo de política pública a facilitação da pesquisa e do desenvolvimento de IA, incentivando a inovação no mercado de novas tecnologias ao invés de impor normas regulatórias mais estritas sobre o uso de IA no país. Evidentemente que, ainda assim, é necessário enfrentar alguns riscos envolvidos no desenvolvimento e na implementação da tecnologia, mas a aposta é que isso

²⁷ NUSSEY, Sam. KELLY, Tim. *Japan leaning towards softer AI rules than EU, official closer to deliberations say*. **Reuters**, 3 de julho de 2023. Acesso em 7 de fevereiro de 2024.

possa ser feito de forma eficiente não a partir de uma legislação específica mas sim a partir de regulações setoriais reativas e flexíveis. Por exemplo, em abril de 2023 o país aprovou uma reforma ao seu Ato de Trânsito para permitir o tráfego de veículos completamente autônomos – que funcionam sem um condutor ou intervenção humana – em suas vias públicas. Nada obstante, a legislação trouxe algumas exigências voltadas à implementação da IA para mitigar os riscos envolvidos em sua implementação²⁸. Ou seja, trata-se de um exemplo de regulação setorial da IA sem a necessidade de uma legislação específica sobre o tema, focando numa abordagem baseada no modelo de incorporação.



4.2. Estados Unidos

Nos Estados Unidos, o principal marco regulatório sobre IA até o momento é uma ordem executiva assinada pelo presidente Joe Biden em outubro de 2023²⁹. Assim como no Japão, não há no país qualquer movimento coordenado em nível nacional para a formulação de uma legislação específica para dispor sobre o desenvolvimento e a implementação de sistemas de IA. A ordem executiva, entretanto, traz alguns dispositivos surpreendentes que merecem destaque: obrigatoriedade dos desenvolvedores das IAs mais avançadas de compartilharem os resultados de testes de segurança e informações críticas com o governo dos EUA, estruturação de um programa nacional de cibersegurança para identificar e corrigir falhas em sistemas de IA, investimento federal no desenvolvimento de IAs que protejam a privacidade de seus usuários, estabelecimento

²⁸ **The Japan Times**. *Fully automated driving under certain conditions allowed in Japan*. 2 de abril de 2023. Acesso em 7 de fevereiro de 2024.

²⁹ KANG, Cecilia. SANGER, David E. *Biden Issues Executive Order to Create A.I. Safeguards*. **The New York Times**, 30 de outubro de 2023. Acesso em 7 de fevereiro de 2024.

de princípios e boas práticas para mitigar os riscos da IA para trabalhadores e, ainda, investimento federal em pesquisa e desenvolvimento tecnológico na área, incluindo programas para atrair talentos estrangeiros.

Ainda em outubro de 2023, o Departamento de Políticas em Ciência e Tecnologia da Casa Branca publicou um documento chamado *Blueprint for an AI Bill of Rights* no qual foram identificados cinco princípios que devem guiar o desenvolvimento e a implementação de sistemas de IA no país. Os princípios apresentados pelo *Blueprint* são: promoção de sistemas de IA seguros e eficientes, proteção contra discriminação algorítmica, proteção de dados e privacidade, notificação ao usuário e explicação de decisões tomadas com base em sistemas de IA e, por fim, possibilidade do usuário requerer a revisão dessas decisões por um humano³⁰. Veja-se, assim, que os EUA adotam, pelo menos até o momento, uma abordagem que mistura elementos do modelo de incorporação com o modelo ético e se baseia em ferramentas de transparência, direitos dos usuários e análise de risco, ainda que não estejam presentes todas as nuances mencionadas acima.



4.3. Reino Unido

Com a saída do Reino Unido da União Europeia em 2020, o panorama de regulação de novas tecnologias no país sofreu mudanças significativas. Isso significa dizer que, por mais que o Reino Unido possa se posicionar como um observador e até mesmo interlocutor dos desdobramentos regulatórios na região, o país precisa debater e

30 Uma explicação mais detalhada dos cinco princípios está disponível em <<https://www.whitehouse.gov/ostp/ai-bill-of-rights>>.

implementar suas próprias iniciativas regulatórias tendo em vista que os diplomas aprovados pela União Europeia não têm mais validade em seu território. No caso da IA, o Reino Unido publicou em 2023 um *policy paper* descrevendo a sua abordagem “pró-inovação” para o desenvolvimento de uma regulação da tecnologia. O documento foi elaborado e apresentado pelo Secretariado de Estado para Ciência, Inovação e Tecnologia³¹. Segundo a Secretária Michelle Donelan, a regulação britânica fará parte de uma rede de programas e investimentos que buscam transformar o Reino Unido em uma “superpotência da IA”, incluindo subsídios para atrair talentos estrangeiros e desenvolver pesquisas acadêmicas de alto impacto no campo.

Assim como o Japão, o Reino Unido coloca a promoção da inovação em primeiro plano, sem desconsiderar a necessidade de enfrentar os riscos apresentados pela tecnologia. Como afirma Donelan, “não pretendemos, inicialmente, introduzir uma nova legislação” tendo em vista que “a pressa para legislar cedo demais pode onerar as empresas de forma indevida”³². Assim sendo, o *policy paper* indica que a melhor abordagem é seguir um modelo baseado em princípios – ou seja, o modelo ético ou principiológico – que possa, de um lado, garantir monitoramento e, do outro, promover colaborações multissetoriais. Isso indica, como visto anteriormente, uma tendência pela implementação de ferramentas de correção. Ademais, a expectativa inicial é que modelo ético seja associado ao modelo de incorporação, ou seja, os princípios irão guiar a implementação de regulações já existentes em diferentes contextos onde a IA já é usada. Segundo o *policy paper*,

31 O *policy paper* com a visão do Reino Unido sobre regulação de IA pode ser acessado em <<https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper>>.

32 *Ibid.*



são, ao todo, cinco princípios que deverão ser considerados para essa tarefa: (1) segurança e robustez, (2) transparência e explicabilidade, (3) justiça, (4) *accountability* e governança e (5) contestação e reparação de danos³³.

4.4. Molduras Transnacionais e Internacionais

Para além das iniciativas nacionais apresentadas acima, é preciso estar atento para algumas molduras transnacionais e internacionais que estão em desenvolvimento e devem influenciar a agenda de regulação da IA pelos próximos anos. Em geral, essas molduras tendem a seguir o modelo ético e, assim, indicam algumas diretrizes de regulação de IA com o intuito de subsidiar debates regulatórios em diferentes contextos. A UNESCO, por exemplo, vem liderando discussões sobre princípios éticos para a regulação da IA desde 2021, quando publicou suas “Recomendações sobre a Ética da Inteligência Artificial” que conta com valores como proteção dos direitos humanos, desenvolvimento sustentável e diversidade e inclusão³⁴. Outro exemplo proeminente são os “Princípios de IA” adotados pela OCDE em 2019, incluindo desenvolvimento inclusivo e sustentável, transparência, explicabilidade e segurança³⁵. Além de princípios, a OCDE também faz recomendações aos reguladores, como investimentos em pesquisa e desenvolvimento, cooperação internacional e formulação de planos de transição para o mercado de trabalho.

³³ *Ibid.*

³⁴ O relatório pode ser acessado em <https://unesdoc.unesco.org/ark:/48223/pf0000381137_por>.

³⁵ Os princípios e recomendações da OCDE estão disponíveis em <<https://oecd.ai/en/ai-principles>>.

4. Conclusão: Regulação da Inteligência Artificial para além da UE

O presente relatório ofereceu um panorama geral sobre os diferentes modelos de regulação de IA, algumas das principais ferramentas à disposição dos reguladores e, ainda, algumas iniciativas regulatórias que divergem da maneira como a União Europeia vem pautando esse debate. A principal conclusão desse estudo preliminar – sem a intenção de esgotar o assunto ou sugerir que outros modelos e ferramentas não existam –, portanto, é que devemos resistir à tentação de assumir um ou outro modelo (e suas respectivas ferramentas) como paradigmático antes que se tenha a chance de testar diferentes soluções e alternativas em contextos distintos. É por isso que, por mais importante que seja o “Efeito Bruxelas” no que diz respeito à apresentação de possíveis molduras regulatórias, seria um equívoco assumir as soluções construídas pela UE como inevitáveis ou até mesmo indispensáveis. É possível – e, mais do que isso, necessário – pensar em abordagens alternativas a partir de novas combinações entre diferentes modelos e ferramentas de regulação da IA.

União Europeia



Acesse nossas redes



itsrio.org