UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
FACULDADE DE CIÊNCIAS HUMANAS E SOCIAIS

GABRIEL AUGUSTO DE PAULA OLIVIERI

GOVERNANÇA GLOBAL DA INTELIGÊNCIA ARTIFICIAL NO COMBATE À DESINFORMAÇÃO

O IMPACTO DO ATO DE IA DA UNIÃO EUROPEIA

Franca
2024

GABRIEL AUGUSTO DE PAULA OLIVIERI

GOVERNANÇA GLOBAL DA INTELIGÊNCIA ARTIFICIAL NO COMBATE À DESINFORMAÇÃO

O IMPACTO DO ATO DE IA DA UNIÃO EUROPEIA

Monografia apresentada à Universidade Estadual Paulista "Júlio de Mesquita Filho" (Unesp), Franca, para obtenção do título de Bacharel em Relações Internacionais.

Orientador: Prof. Dr. Jonathan de Araujo de Assis

Franca

2024

O49g

Olivieri, Gabriel Augusto de Paula
Governança global da inteligência artificial no combate à desinformação : O impacto do Ato de IA da União Europeia / Gabriel Augusto de Paula Olivieri. -- Franca, 2024
51 f.

Trabalho de conclusão de curso (Bacharelado - Relações Internacionais) - Universidade Estadual Paulista (UNESP), Faculdade de Ciências Humanas e Sociais, Franca
Orientador: Jonathan de Araujo de Assis

1. Relações Internacionais. 2. Governança. 3. Inteligência artificial. 4. Desinformação. 5. União Europeia. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Dados fornecidos pelo autor(a).

GABRIEL AUGUSTO DE PAULA OLIVIERI

GOVERNANÇA GLOBAL DA INTELIGÊNCIA ARTIFICIAL NO COMBATE À DESINFORMAÇÃO

O IMPACTO DO ATO DE IA DA UNIÃO EUROPEIA

TCC apresentado à Universidade Estadual Paulista "Júlio de Mesquita Filho" (Unesp), Franca, para obtenção do título de Bacharel em Relações Internacionais.

Banca Examinadora:

_____

Prof. Dr. Jonathan de Araujo de Assis
Instituto de Políticas Públicas e Relações Internacionais (IPPRI-Unesp)

_____

Prof. Dr. Marcelo Passini Mariano
Universidade Estadual Paulista "Júlio de Mesquita Filho" (Unesp) - Campus de Franca

_____

Profa. Ma. Bárbara Campos Diniz
Programa de Pós-Graduação em Relações Internacionais San Tiago Dantas (UNESP, UNICAMP, PUC-SP)

AGRADECIMENTOS

À minha família, por todo o apoio e encorajamento que me deram, o que permitiu que eu me dedicasse aos meus estudos e à pesquisa que resultou neste trabalho.

Ao meu orientador, Jonathan de Araujo de Assis, pela atenção e pela paciência que teve comigo durante a produção deste trabalho.

Aos meus amigos Cássio Aparecido Gimenes Nakashima Júnior, Guilherme Giovaneli Lopes Silva e Henrique Zavalski Mano, que me auxiliaram diretamente com a pesquisa da qual decorreu este trabalho.

RESUMO

Recentemente, houve grande avanço no campo da Inteligência Artificial (IA), levando a um amplo acesso da população a ferramentas que utilizam essa tecnologia. No entanto, com a democratização da IA, decorrem possíveis riscos, como sua utilização para gerar desinformação. Neste trabalho, buscamos responder à seguinte pergunta: qual é o papel da governança global na busca por mitigar os riscos do uso da IA para desinformação? Analisaremos o Ato de IA (AIA), recém adotado no âmbito da União Europeia (UE). Entendemos que o bloco foi a instituição que realizou os avanços mais significativos no âmbito da regulamentação de IA nos últimos anos. Para tanto, inicialmente discutiremos o tema da IA propriamente dito, buscando explicar o que ela é e seu impacto na sociedade, a partir de uma reflexão sobre os possíveis usos de ferramentas de IA para gerar desinformação. Em seguida, pensando na questão da governança global, analisaremos as principais iniciativas e projetos relacionados à regulamentação da IA, buscando compreender as ações que estão sendo tomadas para mitigar os riscos do uso da IA para desinformação. Por fim, será examinado o Ato de IA e os impactos que essa legislação pode ter nas Relações Internacionais (RI).

Palavras-chave: Inteligência Artificial; Desinformação; Deep fake; Governança; União Europeia.

ABSTRACT

Recently, there has been great progress in the field of Artificial Intelligence (AI), leading to widespread access by the population to tools that use this technology. However, with the democratization of AI comes possible risks, such as its use to generate disinformation. In this paper, we seek to answer the following question: what is the role of global governance in seeking to mitigate the risks of the use of AI for disinformation? We will analyze the AI Act (AIA), recently adopted by the European Union (EU). We believe that the bloc has made the most significant advances in AI regulation in recent years. To this end, we will first discuss the topic of AI itself, trying to explain what it is and its impact on society, starting with a reflection on the possible uses of AI tools to generate disinformation. Then, thinking about the question of global governance, we will analyze the main initiatives and projects related to the regulation of AI, seeking to understand the actions being taken to mitigate the risks of the use of AI for disinformation. Finally, we will examine the AI Act and the impacts that this legislation may have on International Relations (IR).

Keywords: Artificial Intelligence; Disinformation; Deep fake; Governance; European Union.

LISTA DE FIGURAS

SUMÁRIO

INTRODUÇÃO

No campo de estudos das Relações Internacionais (RI), muito é discutido sobre tecnologia, especialmente no âmbito da segurança. Nesse contexto, boa parte dos estudos na área tem enfoque nas armas autônomas e suas capacidades bélicas (Chamayou, 2015; Peron, 2016; Heyns, 2016; Garcia, 2018; Asaro, 2019; Bousquet, 2023). Mais recentemente, com o rápido avanço de tecnologias de inteligência artificial (IA), seu potencial para a guerra tem sido explorado amplamente. No entanto, com a crescente democratização de ferramentas de IA ocorrida nos últimos anos, vem sendo exploradas outras funções diversas para essa tecnologia, não necessariamente vinculadas à guerra, mas que têm forte capacidade para gerar conflitos de variadas proporções, dadas as condições corretas. O mais alarmante dessas funções é a criação de *deep fakes*, que são ferramentas que utilizam inteligência artificial para alterar o rosto de indivíduos em vídeos de maneira cada vez mais persuasiva, além de ser uma tecnologia que vem se tornando de mais fácil acesso para a população em geral.

Combinada a más-intenções, essa ferramenta se torna um grande risco à integridade das informações, podendo ser utilizada para disseminar desinformação em larga escala em ambientes virtuais. Na esfera da governança global, a questão da regulamentação de tecnologias de inteligência artificial tem sido progressivamente mais debatida, em um esforço de conter os riscos decorrentes do uso mal-intencionado da IA. Além disso, muitos acadêmicos estão discutindo o tema, com trabalhos publicados nos últimos anos em áreas diversas, como RI (Paterson; Hanley, 2020; Bode, 2024; Roberts et al., 2024) e jornalismo (Prado, 2019). Logo, é possível notar como a questão da IA e sua regulamentação vem crescendo em importância, o que inspirou este trabalho.

Nesse sentido, a pergunta que orienta minha pesquisa é: qual é o papel da governança global na busca por mitigar os riscos do uso da inteligência artificial para desinformação? Para respondê-la, primeiramente, explicaremos a IA em si, em especial buscando entender mais sobre os *deep fakes* e seu impacto na sociedade. Assim, procuraremos refletir sobre os possíveis riscos do uso de ferramentas de IA por atores mal-intencionados. Por fim, buscaremos compreender as ações que estão sendo adotadas no âmbito da governança global para regulamentar o uso da IA, fazendo uma análise das principais iniciativas e projetos propostos sobre o tema da desinformação causada por IA, tanto do setor público quanto do setor privado.

Para atingir esses objetivos, desenvolveremos três capítulos. No capítulo 1, a partir de uma revisão da literatura, apresentaremos as definições de IA, bem como um breve histórico

do desenvolvimento dessa tecnologia e exemplos de suas aplicações, para fins de entendimento geral. Além disso, discutiremos como a IA pode ser usada para disseminar desinformação e como essa desinformação ameaça processos democráticos, que é um importante problema a ser abordado. Por fim, trataremos de exemplos de casos internacionais envolvendo desinformação gerada por IA, para inserir esse debate no campo das Relações Internacionais de maneira mais tangível.

Em seguida, no capítulo 2, com base em uma pesquisa bibliográfica, explicaremos as definições de governança global, sua importância e como ela se insere no tema da IA. Através de um mapeamento, apresentaremos os principais atores e instituições envolvidas no tema, realçando iniciativas e projetos de atores como a Organização das Nações Unidas (ONU) e outras organizações internacionais, além de iniciativas regionais. Além disso, discutiremos a percepção do setor acadêmico sobre as possíveis barreiras à implementação de regulamentações eficazes, que questiona a capacidade da comunidade internacional de trabalhar em conjunto para resolver a questão dos malefícios da IA, além de criticar a grande parte das medidas regulatórias que foram tomadas até o momento.

Por último, no capítulo 3, faremos uma análise mais aprofundada do caso da União Europeia em específico, uma vez que o bloco foi o responsável pelos avanços mais significativos no âmbito da regulamentação da IA até o momento. Para isso, examinaremos o contexto histórico e o desenvolvimento do Ato de IA (AIA), o regulamento europeu de inteligência artificial. Após essa contextualização, comentaremos sobre a estrutura desse regulamento, bem como seus objetivos principais e como ele classifica diferentes tipos de IA com base no risco. Ao final, trataremos do impacto que a abordagem da União Europeia pode trazer para o resto do mundo, visto que outras regiões e países muitas vezes são influenciados pelo bloco europeu, fenômeno conhecido como “Efeito Bruxelas”.

CAPÍTULO 1: INTELIGÊNCIA ARTIFICIAL E DESINFORMAÇÃO: CONCEITOS E IMPACTOS

1.1. INTRODUÇÃO À INTELIGÊNCIA ARTIFICIAL

Antes de tudo, é necessário entender o significado de "inteligência artificial". Em uma pesquisa que analisou 49 documentos de políticas de IA publicados entre 2016 e 2018 por governos nacionais, organizações internacionais, consultorias, think tanks e organizações da sociedade civil na União Europeia e nos Estados Unidos, Ulnicane et al. (2021) observou que, nestes documentos, o termo "inteligência artificial" é, frequentemente, utilizado de maneira abrangente (Ulnicane et al., 2021, p. 159).

Segundo o Comitê Econômico e Social Europeu (CESE), inteligência artificial é um termo amplo que engloba inúmeros subcampos, como computação cognitiva, aprendizado de máquina, inteligência aumentada e IA robótica (European Economic and Social Committee, 2017). Nesse sentido, não há um consenso sobre uma definição precisa para IA. Simon (1995), ao tratar sobre a história da inteligência artificial, comenta que, desde sua criação, a IA tinha como base a capacidade de criar "sistemas que exibissem inteligência, seja como explorações puras sobre a natureza da inteligência, explorações da teoria da inteligência humana ou explorações dos sistemas que poderiam realizar tarefas práticas que exigissem inteligência" (Simon, 1995, p. 96, tradução nossa). Nessa perspectiva, o relatório "Artificial Intelligence and life in 2030" conceitua IA "como um ramo da ciência da computação que estuda as propriedades da inteligência por meio da síntese da inteligência" (Stone et al., 2016, p. 13, tradução nossa).

Já o relatório "Artificial Intelligence in Swedish Business and Society: Analysis of Development and Potential" define IA como a "capacidade de uma máquina de imitar o comportamento humano inteligente" (Vinnova, 2018, p. 6, tradução nossa). Ademais, para a Comissão Europeia (2018), IA diz respeito a "sistemas que exibem comportamento inteligente analisando seu ambiente e realizando ações – com algum grau de autonomia – para atingir objetivos específicos" (European Commission, 2018, p. 1, tradução nossa). Ainda segundo essa definição, os sistemas de IA podem ser baseados em software – por exemplo, assistentes de voz, software de análise de imagens, mecanismos de busca, reconhecimento de fala e reconhecimento facial –, ou podem ser incorporados em dispositivos de hardware – por exemplo, robôs avançados, carros autônomos, drones ou aplicativos da Internet das Coisas (European Commission, 2018, p. 1).

Ainda, Russell e Norvig (2009), elencam oito definições de IA de autores distintos, divididas em quatro abordagens: 1) Pensando humanamente: “o novo e empolgante esforço para fazer os computadores pensarem... máquinas com mentes, no sentido pleno e literal” (Haugeland, 1989), “[a automação de] atividades que associamos ao pensamento humano, atividades como tomada de decisões, solução de problemas, aprendizado...” (Bellman, 1978); 2) Pensando racionalmente: “o estudo das faculdades mentais por meio do uso de modelos computacionais” (Charniak e McDermott, 1985), “o estudo dos cálculos que possibilitam a percepção, o raciocínio e a ação” (Winston, 1992); 3) Agindo humanamente: “a arte de criar máquinas que executam funções que exigem inteligência quando executadas por pessoas” (Kurzweil, 1990), “o estudo de como fazer com que os computadores façam coisas nas quais, no momento, as pessoas são melhores” (Rich e Knight, 1991); 4) Agindo racionalmente: “a inteligência computacional é o estudo do design de agentes inteligentes” (Poole et al., 1998), “a IA… está preocupada com o comportamento inteligente em artefatos” (Nilsson, 1998).

À respeito dessas definições, Russell e Norvig (2009) as classificam em dimensões distintas. As definições 1 e 2 estão relacionadas a processos de pensamento e raciocínio, enquanto as definições 3 e 4 tratam de comportamento. As definições 1 e 3 medem o sucesso em termos de fidelidade ao desempenho humano, enquanto as definições 2 e 4 medem em relação a uma medida de desempenho ideal, chamada racionalidade. Os autores complementam afirmando que “um sistema é racional se ele faz a ‘coisa certa’, considerando o que sabe” (Russell; Norvig, 2009, p. 2, tradução nossa). Por fim, Russell e Norvig (2009) reforçam que todas as abordagens foram adotadas historicamente, por pessoas diferentes com métodos diferentes.

Pesquisas naquilo que atualmente compõem o campo da IA já ocorriam desde 1950. No entanto, o termo “inteligência artificial” viria a ser cunhado apenas em 1956, por John McCarthy, no contexto do Projeto de Pesquisa de Verão de Dartmouth (Vinnova, 2018, p. 25). Entre o final da década de 1950 e meados da década de 1970, a IA prosperou, uma vez que computadores se tornaram mais rápidos, mais baratos e mais acessíveis. Porém, nas décadas de 1990 e 2000, apesar dos desenvolvimentos na área, houve uma diminuição no interesse e investimento da IA, que passou por um período de estagnação (Anyoha, 2017).

Mais recentemente, a partir de finais dos anos 2010, a inteligência artificial voltou a ser discutida com mais intensidade, naquilo que Spencer (2019) chamou de “hype da IA”. De acordo com Sloane (2022), três avanços possibilitaram a nova onda da IA: a disponibilidade de grandes conjuntos de dados, o rápido avanço do maquinário computacional e da capacidade de processamento e a invenção de algoritmos de autoaprendizagem baseados em

redes neurais artificiais – “deep learning” (Sloane, 2022). O tema da IA tornou-se ainda mais público ao final de 2022 e início de 2023, com a comercialização de sistemas de IA como o ChatGPT, desenvolvido pela OpenAI (Roberts et al., 2024).

Como ressalta Sloane (2022), recursos de IA estão presentes em diversos níveis no nosso cotidiano, como filtros de spam de e-mails, navegadores de internet e sites de compra. Ho et al. (2023) destaca usos positivos dessa tecnologia, apontando aspectos econômicos, como a melhoria na produtividade de trabalhadores, além de auxiliar na resolução de questões sociais e tecnológicas. Ademais, a Comissão Europeia (2018) enfatiza o papel da IA em contribuir para alcançar os Objetivos de Desenvolvimento Sustentável das Nações Unidas e enfrentar outros desafios “desde o tratamento de doenças crônicas até a redução das taxas de mortalidade em acidentes de trânsito, o combate às mudanças climáticas ou a antecipação da segurança cibernética” (European Commission, 2018, p. 1, tradução nossa). Por fim, o plano estratégico nacional de pesquisa e desenvolvimento de IA dos Estados Unidos apresenta a IA como “uma tecnologia transformadora que promete um enorme benefício social e econômico. A IA tem o potencial de revolucionar a forma como vivemos, trabalhamos, aprendemos, descobrimos e nos comunicamos” (Conselho Nacional de Ciência e Tecnologia, 2016, p. 3, tradução nossa).

Por outro lado, muitos pesquisadores também frisaram os pontos negativos do uso da IA. De acordo com Ho et al. (2023) “esses sistemas também apresentam desafios, incluindo deslocamento da força de trabalho, falta de transparência, resultados tendenciosos, benefícios compartilhados de forma desigual e ameaças à segurança nacional” (Ho et al., 2023, p. 4, tradução nossa). Roberts et al. (2024) ainda aponta que os sistemas de IA “ameaçam a segurança nacional ao democratizar recursos que podem ser usados por agentes mal-intencionados” (Roberts et al., 2024, p. 1275, tradução nossa).

Assim, é possível observar que a questão da democratização da tecnologia é uma preocupação pertinente, visto seu potencial uso nocivo. Prado (2019) afirma que democratizar a participação da população na internet é, por um lado, algo positivo, mas que também traz problemas ao dar voz a pessoas inexperientes. Dessa maneira, a informação entra em desordem, uma vez que surge uma grande quantidade de notícias fraudulentas, que têm como consequências “induções a erros, má-fé, chance para a patifaria agir, falsificação da realidade para contemplar interesses escusos” (Prado, 2019, p. 2). A autora ainda destaca que, apesar de problemas como esse já existirem há muito tempo, a velocidade com a qual se desenvolveu a tecnologia nos últimos anos contribuiu significativamente para sua intensificação.

1.2. INTERSEÇÃO ENTRE IA E DESINFORMAÇÃO

Considerando os efeitos sociais da democratização de tecnologias de IA, que se manifestam através da internet, muitos utilizam essas mesmas tecnologias para monitorar e regular conteúdos publicados online. Gillespie (2014) comenta sobre os algoritmos de pesquisa que, antigamente, eram baseados apenas em exibir a frequência com a qual certos termos de pesquisa apareciam nas páginas da internet. Atualmente, a esses algoritmos foram adicionadas informações sobre o contexto dos sites. Os algoritmos "consideram a frequência e como o site é relacionado por outros; e convocam técnicas de processamento de linguagem natural para melhor 'entender' tanto a consulta, quanto os recursos que o algoritmo pode oferecer como resposta" (Gillespie, 2014, p. 103). Outro exemplo apontado pelo autor é sobre como o YouTube, que é uma plataforma do Google, "rebaixa algoritmicamente" certos vídeos para que não recebam destaque na plataforma. Também é destacado como exemplo o X – antigo Twitter – que, apesar de não censurar determinados conteúdos, utiliza de seu algoritmo para evitar que tais conteúdos sejam exibidos na aba de assuntos do momento. Dessa forma, complementa Gillespie (2014), fica evidente como a grande mídia, com seus algoritmos baseados em IA, tem o poder de determinar o que será visto e o que será deixado de fora, moldando o discurso público nas redes sociais.

No entanto, a IA também favorece o surgimento de novos problemas, uma vez que facilita a criação de imagens manipuladas, porém convincentes, viabilizando campanhas de desinformação personalizadas, o que as tornam muito mais eficientes (Prado, 2019, p. 5). Nesse contexto, cria-se um enorme potencial de uso de ferramentas de inteligência artificial para criação e disseminação de notícias falsas – ou *fake news*, como são conhecidas –, termo associado à desinformação, que ocorre quando

> informações inventadas para produzir lucros ou comprometer a reputação das pessoas passam a influenciar o debate público nas redes e fora delas. Uma característica central desses conteúdos é que são produzidos de forma organizada e intencional para enganar. A desinformação adquiriu uma velocidade maior na era da Internet, transformando-se em parte do jogo político de nossas sociedades. Ela inclui não só as notícias falsas, mas também a publicação proposital de uma notícia antiga ou fora de contexto e a mobilização de grandes grupos — ou até mesmo robôs — para reforçar determinados discursos (Comitê Gestor da Internet no Brasil, 2018, p. 38).

No entanto, o termo vem sendo usado com menos frequência, uma vez que foi taxado como inadequado por muitos autores. Wardle e Derakhshan (2017) afirmam que o termo "notícias falsas" não é capaz de "descrever os fenômenos complexos da poluição de informações" (Wardle; Derakhshan, 2017, p. 5, tradução nossa). Os autores complementam ao

reiterar como o termo vem sendo utilizado por políticos para descrever qualquer coisa que lhes é desagradável, apontando como "ele está se tornando um mecanismo pelo qual os poderosos podem reprimir, restringir, minar e contornar a imprensa livre" (Wardle; Derakhshan, 2017, p. 5, tradução nossa). Nesse sentido, os autores apresentam três conceitos para analisar o distúrbio de informações: *mis-information* – quando informações falsas são compartilhadas, mas sem intenção de causar danos –, *dis-information* – quando informações falsas são compartilhadas conscientemente para causar danos – e *mal-information* – quando uma informação genuína é compartilhada para causar danos, geralmente movendo para a esfera pública informações destinadas a permanecer privadas (Wardle; Derakhshan, 2017, p. 5). No contexto da *dis-information*, ou desinformação, Wardle e Derakhshan (2017) prevêem um aumento no uso de tecnologias de inteligência artificial tanto para criar, quanto para identificar desinformações.

Em meio à essa crescente desinformação, uma tecnologia que vem recebendo destaque é o chamado *deep fake*. Essa nova ferramenta utiliza inteligência artificial para trocar o rosto de pessoas em vídeos de maneira extremamente persuasiva, sincronizando até mesmo expressões faciais e o movimento dos lábios durante a fala. Para chegar a esse resultado, a ferramenta precisa apenas de um vídeo com o rosto da vítima para que possa mapeá-lo e colocá-lo no vídeo final (Shimabukuro; Marques, 2017). Chesney e Citron (2018) explicam como os rápidos avanços em algoritmos de *deep learning* possibilitaram a criação de *deep fakes* cada vez mais realistas e difíceis de serem detectados. Os autores complementam que "com a disseminação dessa tecnologia, a capacidade de produzir conteúdo de vídeo e áudio falso, porém acreditável, estará ao alcance de uma gama cada vez maior de governos, agentes não estatais e indivíduos" (Chesney; Citron, 2018, tradução nossa).

Assim, é esperado que as capacidades de criar conteúdos falsos se espalhem cada vez mais, à medida que seus métodos se aperfeiçoam. Chesney e Citron (2018) mostram, ainda, como o atual ambiente onde são compartilhadas informações é propício para a disseminação de mentiras. Segundo pesquisa realizada pelo Pew Research Center em 2017, 67% dos estadunidenses afirmaram que recebem pelo menos parte das notícias por meio de mídias sociais (Shearer; Gottfried, 2017). Segundo Chesney e Citron (2018) isso é resultado do declínio da credibilidade e influência dos grandes canais de mídia e do aumento da importância das redes sociais para muitas pessoas, o que as torna expostas à todo tipo de desinformação, uma vez que essas mídias são uma fonte de notícias comparativamente não filtradas. Dessa forma, afirmam os autores, foi criado um terreno fértil para a circulação de conteúdo falso.

Naturalmente, muitos pesquisadores expressam preocupação com esses desenvolvimentos. Paterson e Hanley (2020) indicam como as novas ferramentas de *deep fake* tem enorme potencial para aumentar a eficácia de campanhas de guerra política. Em consequência, campanhas de guerra política bem-sucedidas podem se tornar ameaças aos Estados democráticos, abalando suas estruturas basilares e se tornando um risco existencial. Os autores afirmam que "uma 'deep fake' bem cronometrada e disseminada, ou uma série de 'deep fakes', poderia ser usada com eficácia para enganar os eleitores de forma abrangente e criar ou exacerbar tensões políticas, influenciando fortemente e talvez até mesmo alterando eleições democráticas" (Paterson e Hanley, 2020, p. 448, tradução nossa). Tal preocupação é também expressada por Chesney e Citron (2018) e Polyakova e Boyer (2018).

1.3. IA E DESINFORMAÇÃO NAS RELAÇÕES INTERNACIONAIS

Em relação aos estudos de RI sobre tecnologias de IA, Bode (2024) comenta como a literatura se concentra principalmente em analisar campanhas de desinformação provenientes de Estados influentes, como por exemplo a Rússia e a China – e até mesmo de atores não-estatais –, em contextos de guerra híbrida, que combina aspectos tradicionais e cibernéticos. Nesse sentido, Gerrits (2018) observa como a desinformação difundida por atores estatais e não estatais expressa "uma grande ameaça às democracias ocidentais e às instituições internacionais que elas construíram" (Gerrits, 2018, p. 6, tradução nossa). Logo, complementa Bode (2024), "os estudos de RI [...] tendem a ser de natureza interdisciplinar, baseando-se, por exemplo, significativamente em percepções dos estudos de mídia e comunicação" (Bode, 2024, p. 6, tradução nossa).

No âmbito da proliferação de *deep fakes*, já é possível observar alguns casos de seu uso na mídia[1]. Um exemplo de uso positivo dessa tecnologia é evidenciado no documentário *Welcome to Chechnya*, dirigido por David France e estreado na HBO em 2020. No documentário, são relatadas experiências de pessoas LGBTQ que sofreram perseguições na Chechênia, região da Rússia de maioria muçulmana. Para evitar consequências negativas para os entrevistados, France optou por utilizar *deep fakes* para alterar digitalmente seus rostos, tomando como modelo os rostos de voluntários da comunidade LGBTQ. O diretor do documentário defendeu que, apesar de muitas reportagens sensíveis fazerem uso de outros

[1] Para mais conteúdo sobre *deep fakes*, assista ao vídeo "I Deep Faked Myself, Here's Why It Matters". JOHNNY HARRIS. I Deep Faked Myself, Here's Why It Matters. [vídeo]. YouTube, 19 jul. 2023. Disponível em: <https://www.youtube.com/watch?v=S951cdansBI>. Acesso em: 15 nov. 2024.

tipos de censura para manter o anonimato dos entrevistados, como colocá-los em uma sombra ou desfocar seus rostos, isso diminui a imersão do programa. France comenta que o *deep fake* "permite que meus participantes narrem suas próprias histórias. E isso lhes devolve a humanidade de uma forma que não seria possível em outras circunstâncias" (Rothkopf, 2020, tradução nossa).

Surgiram também diversos casos de atores fazendo uso de ferramentas de IA para produzir *deep fakes* de figuras públicas. Em 2017, pesquisadores da Universidade de Washington utilizaram inteligência artificial para produzir um vídeo do ex-presidente dos Estados Unidos, Barack Obama, no qual ele dizia coisas que de fato havia dito, porém em um contexto diferente. O vídeo necessitou de várias horas de gravações do ex-presidente, para que então fosse possível treinar o algoritmo para replicar seu rosto e ajustar o movimento dos lábios no vídeo criado (Suwajanakorn et al., 2017). Em 2019, havia preocupações nos Estados Unidos sobre o possível uso de *deep fakes* para influenciar as eleições presidenciais de 2020. No entanto, segundo Siwen Lyu – professor do Departamento de Ciência e Engenharia da Computação da Universidade Estadual de Nova York em Buffalo – isso não ocorreu pois a tecnologia ainda não havia sido suficientemente desenvolvida e disseminada (Metz, 2022).

Em alguns casos, vídeos editados foram utilizados de maneira mal-intencionada para atacar figuras públicas. Um exemplo disso ocorreu em 2019, quando circulou nas redes sociais um vídeo da ex-presidente da Câmara dos Representantes dos Estados Unidos, Nancy Pelosi, no qual a democrata aparentava estar bêbada e com dificuldades na fala. Segundo Budryk (2019), o vídeo viralizou em um momento onde muitos especialistas estavam preocupados com o uso de vídeos editados para espalhar desinformação, principalmente utilizando ferramentas como *deep fakes*. Apesar de não fazer uso dessa ferramenta, o vídeo levantou questões sobre como a desinformação pode ser utilizada como arma política, como o que foi proposto também por Paterson e Hanley (2020) ao comentarem sobre as ameaças que a desinformação pode trazer para eleições. Hany Farid, professor de ciência da computação e especialista em perícia digital da Universidade da Califórnia, afirmou em uma entrevista ao Washington Post que "é impressionante que uma manipulação tão simples possa ser tão eficaz e acreditável para algumas pessoas" (Harwell, 2019, tradução nossa).

Já no contexto da Guerra Russo-Ucraniana, em março de 2022, poucas semanas após a deflagração do conflito, foi divulgado nas redes sociais um suposto vídeo do presidente ucraniano Volodymyr Zelensky, no qual ele exigia que seus soldados se rendessem. O vídeo foi ridicularizado pelos ucranianos, uma vez que a qualidade do *deep fake* não era muito alta, tornando claro que se tratava de um vídeo manipulado. O vídeo foi rapidamente tirado do ar

em diversas plataformas, mas não antes de o verdadeiro presidente Zelensky divulgar um vídeo no Instagram denunciando o que ele chamou de “provocação infantil” (Wakefield, 2022). Nina Schick, autora do livro “Deepfakes: The Coming Infocalypse”, comenta como o vídeo era medíocre e facilmente detectado como falso até mesmo por “espectadores semissofisticados”. A autora argumenta que “mesmo que esse vídeo tenha sido muito ruim e grosseiro, esse não será o caso em um futuro próximo. [...] Pode corroer a confiança na mídia autêntica” (Wakefield, 2022, tradução nossa).

Ademais, Schick complementa afirmando que “as pessoas começam a acreditar que tudo pode ser falsificado” (Wakefield, 2022), e que essa tecnologia “é uma nova arma e uma forma potente de desinformação visual – e qualquer um pode fazer isso” (Wakefield, 2022). Ainda no contexto da guerra, o lado russo também foi alvo de desinformação. Em junho de 2023, foi ao ar em várias redes de rádio e televisão russas um suposto discurso do presidente Vladimir Putin, no qual ele pedia que cidadãos russos fugissem para o interior do país por conta de uma invasão ucraniana a três regiões fronteiriças (Sonne, 2023). O porta-voz do Kremlin, Dmitry Peskov, descreveu o incidente como um “hack” e que o presidente Putin não havia gravado nenhum discurso de emergência (Sonne, 2023).

Atualmente, está cada vez mais fácil fazer uso de ferramentas de *deep fake* (Metz, 2022). Ao falar sobre a Guerra Russo-Ucraniana, Dan Coats – ex-diretor de inteligência nacional dos EUA – e Siwei Lyu afirmaram como o fato de *deep fakes* estarem sendo usados na tentativa de influenciar as pessoas durante uma guerra é perigoso, uma vez que a confusão que eles criam pode ser grave. Lyu complementa que, em casos normais, o uso de um *deep fake* pode gerar curiosidade na internet, mas não necessariamente causar danos. No entanto, “em situações críticas, durante uma guerra ou um desastre nacional, quando as pessoas realmente não conseguem pensar de forma muito racional e têm apenas um espaço de atenção muito curto, e veem algo assim, é quando isso se torna um problema” (Metz, 2022, tradução nossa).

É curioso pensar que, há alguns anos, não se falava em *deep fake* e, atualmente, eles estão sendo utilizados até mesmo para influenciar o curso de uma guerra (Metz, 2022). Assim, em um cenário de rápido desenvolvimento e democratização dessa tecnologia, não faltam preocupações envolvendo seu possível uso mal-intencionado, uma vez que se trata de uma tecnologia potencialmente nociva. Logo, a questão da regulamentação veio à tona, impulsionando atores internacionais, que já vinham desenvolvendo estratégias nacionais de IA, a acelerarem esse processo.

CAPÍTULO 2: GOVERNANÇA GLOBAL E A REGULAMENTAÇÃO DA INTELIGÊNCIA ARTIFICIAL

2.1. INTRODUÇÃO À GOVERNANÇA GLOBAL

A governança é um tema de especial importância no âmbito das tecnologias emergentes, uma vez que estas tecnologias trazem grande impacto para a sociedade. No entanto, o termo "governança" apresenta diversos significados, a depender da literatura e contexto no qual está inserido. Nesse sentido, buscando refletir sobre o tema, Ulnicane et al. (2022) fornece percepções sobre os estudos e conceitos de governança, apresentando e comentando definições de diferentes autores.

Segundo Chhotray e Stoker (2009), governança diz respeito às "regras de tomada de decisões coletivas em ambientes onde há uma pluralidade de atores ou organizações e onde nenhum sistema de controle formal pode ditar os termos da relação entre esses atores e organizações" (Chhotray; Stoker, 2009, p. 3, tradução nossa). Baseando-se nessa definição, a governança incluiria tanto regras formais quanto informais, seria composto por decisões tomadas por um conjunto de indivíduos sobre questões de influência e controle mútuos, além de "um amplo entendimento da tomada de decisões que pode ser estratégica, mas também pode estar contida na prática de implementação diária de um sistema ou organização" (Ulnicane et al., 2022, p. 34, tradução nossa).

Em outra perspectiva, Borrás e Edler (2014) definem governança como "os mecanismos pelos quais os atores sociais e os atores estatais interagem e se coordenam para regular questões de interesse social" (Borrás; Edler, 2014, p. 13-14, tradução nossa). Segundo essa perspectiva de governança, complementa Ulnicane et al. (2022), o Estado não atua sozinho, mas coordena suas ações com uma série de outros atores, que incluem o setor privado, a sociedade civil e comunidades especialistas.

Segundo Barnett, Pevehouse e Raustiala (2021), no campo das RI, a governança sofreu alterações nas últimas décadas. Os autores identificam que essas mudanças foram estimuladas por nove grandes fatores estruturais, sendo eles:

> (1) mudanças geopolíticas, como o declínio relativo do poder dos Estados Unidos e a ascensão da China; (2) mudanças na economia global; a grande "aglomeração" no espaço de governança global, tanto em termos de (3) número de atores quanto de (4) pluralização de atores; (5) a complexidade cada vez maior dos problemas globais; (6) mudanças na ideologia e tendências na teoria da governança; (7) a mudança global para a especialização como forma de racionalizar a governança; (8) mudanças tecnológicas; e (9) mudanças políticas internas, por exemplo, na forma de populismo

> e nacionalismo crescentes (Barnett; Pevehouse; Raustiala, 2021, p. 6, tradução nossa).

Em complemento, os autores afirmam que esses fatores são interdependentes, citando como exemplo o fato de que a evolução tecnológica permitiu o surgimento de atores não estatais na governança global. Assim, reforçam, fica claro que houve uma diversificação dos participantes da governança global (Barnett; Pevehouse; Raustiala, 2021).

Já no âmbito da inteligência artificial, o termo "governança" é recorrente em documentos oficiais de planos de IA desenvolvidos por Estados. O termo é comumente utilizado no contexto do governo, da regulamentação e da ética. No entanto, costuma ser vago, sem receber uma definição precisa. Apesar disso, segundo Ulnicane et al. (2022), é possível notar que o termo está intimamente relacionado às discussões sobre política de IA. Nesse contexto, autores como Drezner (2019), Milner e Solstad (2021), e Büthe et al. (2022) defendem a importância da governança de IA, uma vez que, sem interferência política, tendo em vista que essa tecnologia está se espalhando aceleradamente, ela pode afetar estruturas políticas e econômicas nacionais ao redor do mundo.

Nessa conjuntura, Bode (2024) comenta que, no campo das RI, o tema da governança de IA vêm sendo frequentemente discutido, reunindo setores acadêmicos e juristas, para refletir sobre como a IA pode ser governada, os possíveis obstáculos para esta governança e o que poderia acontecer sem uma governança efetiva. Além disso, a autora menciona que parte da literatura de RI que trabalha temas de inteligência artificial se concentra em resumir e categorizar iniciativas de governança de IA. Tais iniciativas possuem naturezas distintas, sendo algumas formadas por entes estatais e outras provenientes da colaboração entre atores do setor privado, empresas de diversas áreas e até mesmo universidades. No entanto, apesar destas diferenças, é possível notar que essas iniciativas buscam um mesmo objetivo: pensar e desenvolver uma base legal que regule e garanta que ferramentas de inteligência artificial sejam implementadas e utilizadas de maneira consciente, segura e confiável.

Dentre os principais atores mundiais que vêm desenvolvendo estratégias de IA, é possível destacar a Organização das Nações Unidas (ONU), a Organização para a Cooperação e Desenvolvimento Econômico (OCDE), grupos como o G7, o BRICS, organizações de caráter regional como a União Europeia (UE), o Mercosul, a União Africana, a Associação de Nações do Sudeste Asiático (ASEAN), além de outras parcerias. Nesse cenário, em meio às inúmeras preocupações decorrentes do avanço da tecnologia de inteligência artificial, percebe-se que a questão da desinformação é uma das mais imediatas. Assim, é importante

destacar o que as principais iniciativas globais de IA vêm propondo no que diz respeito ao uso dessa tecnologia no campo informacional.

2.2. DESAFIOS E OPORTUNIDADES NA REGULAMENTAÇÃO DA INTELIGÊNCIA ARTIFICIAL

Nesse ambiente de surgimento e evolução de inúmeras iniciativas de inteligência artificial, Bode (2024) reconhece que há uma clara percepção de que a governança global de IA é necessária. Contudo, complementa Bode (2024), o rápido crescimento no número de iniciativas de IA nos últimos anos tornou a difícil acompanhar todas elas.

Seguindo essa perspectiva, Roberts et al. (2024) argumenta que o número de novas instituições internacionais no âmbito da IA "exacerbou a fragmentação institucional e a sobreposição de mandatos, limitando a eficácia dos regimes" (Roberts et al., 2024, p. 1280, tradução nossa). Segundo o autor, essa fragmentação do cenário internacional intensifica problemas de cooperação entre atores, uma vez que estes podem seguir normas e padrões diferentes no que concerne à IA.

Ainda destacando as dificuldades de uma governança efetiva de IA, Büthe et al. (2022) define a IA como um "alvo móvel", ou seja, em constante evolução, o que torna sua governança um enorme desafio para instituições públicas e democráticas. Ainda mais, o autor compara as tentativas de governar a inteligência artificial com uma tentativa de "pastorear gatos". O autor aponta que a IA é uma tecnologia em rápida mudança, que está inclinada a se desenvolver e se adaptar por conta própria em um ritmo capaz de superar a capacidade humana. Logo, "se a aprendizagem autônoma é uma característica definidora (mesmo que ainda bastante limitada) da IA, então governar a IA é uma tentativa de regular algo com uma (mais ou menos proeminente) mente própria" (Büthe et al., 2022, p. 1733, tradução nossa).

Por fim, alguns autores questionam a capacidade das instituições já existentes de enfrentar os problemas decorrentes do desenvolvimento da inteligência artificial. Bode (2024) aponta como muitos Estados e organizações julgam a legislação atual como insuficiente para responder aos novos desafios proporcionados pela IA. Do mesmo modo, Taeihagh (2021) comenta que as estruturas regulatórias e de governança existentes não possuem o aparato necessário para lidar com as questões trazidas pela IA por conta da "insuficiência de informações necessárias para entender a tecnologia e as defasagens regulatórias por trás dos desenvolvimentos da IA" (Taeihagh, 2021, p. 144, tradução nossa). O autor complementa afirmando que as *big techs* possuem vantagens em termos de informações e recursos quando

comparados a governos. Nesse sentido, as iniciativas de IA que combinam tanto atores estatais quanto empresas de tecnologia podem se colocar como um meio capaz de solucionar essas disparidades.

Nesse cenário, Bode (2024) também indica a existência de uma "lacuna de governança internacional nas tecnologias de IA" (Bode, 2024, p. 5, tradução nossa). Segundo a autora, enquanto não houver esforços mais expressivos no sentido de uma forma mais ampla e preparada de governança internacional de IA, essa lacuna continuará a existir. Seguindo essa ideia, Roberts et al. (2024) defende que "fortalecer o complexo regime de IA existente em vez de desenvolver novas instituições centralizadas é a opção de governança mais desejável e realista" (Roberts et al., 2024, p. 1284, tradução nossa).

Por outro lado, Kuhlmann et al. (2019), comenta sobre a governança provisória, conceito que pode ser entendido como uma abordagem mais cautelosa ao estabelecer normas sobre tecnologias emergentes. Seguindo esse conceito, busca-se estabelecer, primeiramente, um *framework* em uma certa direção, sem regras muito definitivas. Dessa forma, é possível observar como a situação se desenvolve e adaptar as decisões tomadas em relação a esses desenvolvimentos. De acordo com o autor, a governança provisória se mostra prudente no campo de novas tecnologias, como o da IA, uma vez que "as incertezas da evolução da ciência e da tecnologia colocam desafios específicos à governança destes campos, que têm de lidar com objetivos mal definidos e muitas vezes 'alvos móveis'" (Kuhlmann et al., 2019, p. 1091, tradução nossa).

Até o momento, como explicita Bode (2024), as iniciativas de governança de IA que obtiveram algum sucesso se concentraram em elaborar declarações ou listas de princípios, ou seja, atuam como uma lei internacional branda. Contudo, para que seja possível pensar em soluções globais para a governança de IA, é proveitoso observar iniciativas que já deram um passo além das outras, em direção à formulação de leis de caráter obrigatório. Nesse sentido, torna-se relevante analisar os projetos mais sólidos até o momento, para que deles se possa extrair lições e melhores práticas para lidar com os novos desafios da IA.

2.3. INICIATIVAS E PROJETOS PARA COMBATER A DESINFORMAÇÃO GERADA POR INTELIGÊNCIA ARTIFICIAL

Em relação à Organização das Nações Unidas (ONU), foi criado em 2017 a *AI for Good*, uma iniciativa vinculada à União Internacional de Telecomunicações (ITU). Através de cúpulas anuais, a *AI for Good Global Summit*, são promovidos eventos para se pensar e

discutir soluções de IA que estejam de acordo com as metas de desenvolvimento sustentável estabelecidas pela ONU. Na cúpula de 2024, foi tratado, dentre outros temas, sobre a questão de IA generativa e *deep fakes*, buscando pensar padrões que garantam a autenticidade das informações divulgadas na Internet. Ainda no âmbito da ONU, foi adotada pela Assembléia Geral, em março de 2024, a primeira resolução sobre a regulamentação da IA[2]. O texto trata sobre a promoção de sistemas de inteligência artificial "seguros, protegidos e confiáveis" e que estejam de acordo com o desenvolvimento sustentável. Além disso, a Assembleia solicitou a todos os Estados, o setor privado, a sociedade civil, as organizações de pesquisa e a mídia que buscassem elaborar e apoiar abordagens e diretrizes regulatórias e de governança referentes à utilização assegurada e transparente da IA (Mishra, 2024, tradução nossa).

Ainda no âmbito da ONU, em seu discurso na abertura da 79ª Assembléia Geral, o presidente Lula advogou em favor de uma governança intergovernamental de IA:

> Na área de Inteligência Artificial, vivenciamos a consolidação de assimetrias que levam a um verdadeiro oligopólio do saber. Avança a concentração sem precedentes nas mãos de um pequeno número de pessoas e de empresas, sediadas em um número ainda menor de países. Interessa-nos uma Inteligência Artificial emancipadora, que também tenha a cara do Sul Global e que fortaleça a diversidade cultural. Que respeite os direitos humanos, proteja dados pessoais e promova a integridade da informação. E, sobretudo, que seja ferramenta para a paz, não para a guerra. Necessitamos de uma governança intergovernamental da inteligência artificial, em que todos os Estados tenham assento (Brasil, 2024a).

Continuando na esfera de normas propriamente ditas, o Conselho da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) adotou em 2019 a primeira norma intergovernamental sobre IA. A Recomendação sobre Inteligência Artificial[3] tem como objetivo promover a inovação e a confiança na IA, garantindo o respeito aos direitos humanos e valores democráticos. Seus princípios incluem crescimento inclusivo, respeito ao estado de direito, transparência, robustez e segurança, e responsabilidade. Mais tarde, em 2023 e 2024, a Recomendação foi revisada e atualizada para refletir avanços tecnológicos, incluindo IA generativa, e abordar questões como desinformação e sustentabilidade ambiental.

No contexto de blocos regionais, também é possível observar o desenvolvimento de estratégias de IA. No caso da União Africana, em junho de 2024, mais de 130 ministros e especialistas em tecnologia e comunicação do continente se reuniram virtualmente para a 2ª Sessão Extraordinária do Comitê Técnico Especializado em Comunicação e TIC (Tecnologias

---

[2] *Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development: resolution / adopted by the General Assembly*. Disponível em: <https://documents.un.org/doc/undoc/ltd/n24/065/92/pdf/n2406592.pdf>. Acesso em: 28 out. 2024.

[3] *Recommendation of the Council on Artificial Intelligence*. Disponível em: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>. Acesso em 15 set. 2024.

da Informação e Comunicação). Nessa circunstância, foi aprovada a Estratégia Continental de Inteligência Artificial[4] e o Pacto Digital Africano[5], projetos ligados à Estratégia de Transformação Digital da União Africana (2020-2030) e na Agenda 2063, baseadas em políticas como a promoção do desenvolvimento digital e segurança cibernética no continente. Nesse sentido, a Estratégia Continental de IA fornece um guia para os países africanos para aproveitar os benefícios da inteligência artificial, utilizá-la para desenvolvimento da África e prosperidade da população, além de promover o uso ético e buscar minimizar os riscos.

Já no sudeste asiático, em dezembro de 2023, a ASEAN iniciou discussões sobre uma política de inteligência artificial. Em um workshop online, foram reunidas autoridades de diferentes setores da ASEAN, além de representantes de mais de 70 países. O objetivo do evento foi proporcionar um espaço para discussões sobre estratégias de desenvolvimento de inteligência artificial de maneira ética. Em 2024, a ASEAN publicou o *ASEAN Guide on AI Governance and Ethics*[6], um guia onde são oferecidas sugestões de conduta para organizações que queiram desenvolver sistemas de IA na região. No documento, é destacada a necessidade de aumentar o conhecimento da população sobre os potenciais riscos e benefícios da inteligência artificial para que as pessoas possam se proteger de usos nocivos de sistemas de IA (ASEAN Secretariat, 2024). Assim, para que se eleve o nível de conscientização público, é sugerido aumentar o contato das pessoas com sistemas de inteligência artificial em seu cotidiano. Segundo o documento, isso pode ser feito por meio de redes sociais, fóruns públicos, implementação de aplicativos de IA (como chatbots ou assistentes virtuais) e até mesmo investimento em educação. Dessa maneira, estando em contato e entendendo como funciona a inteligência artificial, é esperado que as pessoas aprendam a lidar melhor com essa tecnologia.

No âmbito do Mercosul, em novembro de 2023, foi realizada em Brasília a 42ª Reunião de Altas Autoridades sobre Direitos Humanos do Mercosul (RAADH), onde foi aprovada a Declaração de Princípios de Direitos Humanos no âmbito da Inteligência Artificial no Mercosul[7]. O texto da declaração, que conta com 17 princípios, busca adotar diretrizes com o olhar dos direitos humanos. Alguns dos tópicos mais relevantes do documento incluem

[4] *Continental Artificial Intelligence Strategy*. Disponível em: <https://au.int/sites/default/files/documents/44004-doc-EN-_Continental_AI_Strategy_July_2024.pdf>. Acesso em: 28 out. 2024.
[5] *African Digital Compact*. Disponível em: <https://au.int/sites/default/files/documents/44005-doc-AU_Digital_Compact_V4.pdf>. Acesso em: 28 out. 2024.
[6] *ASEAN Guide on AI Governance and Ethics*. Disponível em: <https://asean.org/book/asean-guide-on-ai-governance-and-ethics/>. Acesso em: 28 jun. 2024.
[7] Declaração de Princípios de Direito Humanos no âmbito da inteligência artificial no MERCOSUL. Disponível em: <https://documentos.mercosur.int/public/declaraciones/165>. Acesso em: 30 out. 2024.

o desenvolvimento responsável da inteligência artificial, a garantia de igualdade e não discriminação, a promoção da educação e formação digital, a eliminação de segmentos étnico-raciais na inteligência artificial e a utilização de tecnologias de reconhecimento facial com transparência, privacidade e sem segmentos raciais. Além disso, durante a LXIII cúpula de presidentes do Mercosul, em dezembro de 2023, os presidentes da Argentina, Brasil, Paraguai, Uruguai, e outras autoridades declararam sobre os riscos impostos pela IA no âmbito da comunicação. Foram destacados

> os desafios gerados pela acelerada transformação tecnológica e os recentes avanços na área da inteligência artificial e, particularmente, a ameaça que a desinformação, os discursos de ódio e a apologia à violência, disseminados em longa escala e alta velocidade pelo fenômeno da "viralização" de conteúdo, representam à coesão social, aos valores e instituições democráticas, aos direitos humanos, à legitimidade do conhecimento científico e à confiança no jornalismo, potencialmente comprometendo o debate público sobre questões relevantes de alcance local e mundial, como a mudança do clima ou a pandemia de coronavírus, independentemente da consciência, intencionalidade ou motivação de quem os propaga. Por fim, solicitaram às instâncias pertinentes no Mercosul a refletir sobre o conteúdo da presente Declaração em suas atividades (Mercosul, 2023).

Ademais, foi proposto o Fórum Internacional "Crimes virtuais e desinformação política: ameaças à integridade eleitoral", um espaço para dialogar a respeito dos avanços tecnológicos e seu uso tanto positivo quanto negativo para a democracia, com o intuito de aprofundar o debate sobre o tema. O fórum ocorreu no dia 04 de julho de 2024 no Auditório do Tribunal Superior de Justiça Eleitoral (TSJE) em Assunção, Paraguai, e contou com a participação de autoridades do Mercosul, representantes de organizações internacionais, atores do setor privado e do setor público, além do meio acadêmico.

Já no caso dos BRICS, o grupo demonstrou interesse em se tornar um líder no campo da inteligência artificial, uma vez que o assunto vem sendo discutido há algum tempo pelos países membros do grupo. Em uma reunião com o chanceler da Rússia, Sergey Lavrov, ocorrida em fevereiro de 2024, o presidente Lula reafirmou a importância de "uma nova governança global para lidar com temas como inteligência artificial e as mudanças climáticas, que não podem ser enfrentadas isoladamente" (Brasil, 2024b). Especialistas destacam, no entanto, os desafios que o BRICS deve enfrentar para alcançar uma posição central do debate global de IA, como a dificuldade de estabelecer um consenso comum entre todos os membros do bloco, principalmente após sua recente expansão.

Adicionalmente, em 2023, foi lançado pelo G7 o *Hiroshima AI Process*, uma iniciativa que possui o intuito de promover o desenvolvimento de inteligência artificial segura e confiável. O grupo de países demonstrou inúmeras preocupações a respeito do rápido avanço tecnológico na área de IA, sendo notável o potencial dessa tecnologia de influenciar

audiências na Internet. Acrescentando a esse problema, foi levantado o tema do uso de IA para criar conteúdos como *deep fakes*, o que poderia elevar o nível de desinformação online e causar riscos à estabilidade social. Para enfrentar esse problema, os membros do G7 argumentam a favor da cooperação global em um investimento em educação com enfoque em "alfabetização digital", uma vez que se tornou difícil para a população acompanhar o rápido desenvolvimento da inteligência artificial.

Do mesmo modo, foi elaborada, em 2020, a *Global Partnership on Artificial Intelligence* (GPAI), uma iniciativa multilateral, até o momento composta por 28 países e pela UE, que visa "preencher a lacuna entre a teoria e a prática em IA, apoiando investigação de ponta e atividades aplicadas em prioridades relacionadas com a IA" (Global Partnership on Artificial Intelligence, 2020, tradução nossa). Seus projetos são divididos em 4 grupos: IA responsável; Gestão de dados; Futuro do trabalho e Inovação e comercialização. No que concerne ao uso de inteligência artificial no setor da informação, o grupo de IA responsável vêm trabalhando em projetos que envolvem questões de regulação de uso em redes sociais, como o *Responsible AI for social media governance*, que defende a necessidade de implementação de mecanismos de detecção de IA, como marcas d'água, por exemplo, para que seja fácil distinguir imagens alteradas por IA. O grupo propõe que qualquer desenvolvedor de modelos de IA deve ser requerido por lei a "demonstrar uma ferramenta de detecção confiável para o conteúdo gerado pelo modelo, como condição para seu lançamento para o público" (Global Partnership on Artificial Intelligence, 2020, p. 8 tradução nossa), visão reforçada por Knott, Pedreschi, Chatila et al. (2023), acadêmicos envolvidos com o projeto. Ainda, segundo o GPAI, essa ferramenta de detecção deve ser de livre acesso para o público.

Similarmente, foi anunciada em 2016 a *Partnership on AI* (PAI), uma parceria entre agentes da sociedade civil, empresas privadas e universidades. Seu trabalho principal consiste em desenvolver ferramentas que auxiliem na divulgação de conhecimento acerca de inteligência artificial, buscando criar um maior preparo da sociedade para lidar com questões referentes a essa tecnologia. Através do programa de *AI and Media Integrity Work* (uma das diversas frentes do trabalho realizado), o grupo desenvolve 4 projetos: Conteúdo Sintético e Manipulado; Segmentação e classificação de conteúdo; Explicações para o público e Notícias locais. Cada um desses projetos explora questões diferentes acerca do uso de IA na Internet, sendo seu trabalho divulgado através de blogs disponíveis no site da organização. Apesar de não possuir fins lucrativos, a *Partnership on AI* conta com uma ampla parceria com atores influentes, notavelmente com as *big techs* Apple, Amazon, Meta, Google/DeepMind, IBM e

Microsoft, as quais tiveram papel fundamental no apoio original do projeto. Uma síntese das iniciativas é apresentada no Apêndice A.

No entanto, dentre as instituições e projetos examinados, a União Europeia é quem realizou avanços mais significativos no âmbito da regulamentação de IA. Recentemente, o bloco aprovou a implementação do Ato de IA, uma medida para controlar a inteligência artificial mais estrita do que as iniciativas observadas até o momento. O capítulo 3 deste trabalho será dedicado a explorar com mais profundidade o caso europeu.

CAPÍTULO 3: A UNIÃO EUROPEIA E A REGULAMENTAÇÃO DA INTELIGÊNCIA ARTIFICIAL

3.1. CONTEXTO HISTÓRICO E DESENVOLVIMENTO DO ATO DE IA

O caso europeu é um importante exemplo no âmbito da governança de inteligência artificial. Em outubro de 2020, o Parlamento Europeu adotou três relatórios que detalham como a UE pode regulamentar a IA de maneira efetiva, ao mesmo tempo em que incentiva a inovação, a ética e a confiança na tecnologia. O Parlamento destaca como um dos relatórios "se concentra em como garantir a segurança, a transparência e a responsabilidade, evitar preconceitos e discriminação, promover a responsabilidade social e ambiental e garantir o respeito aos direitos fundamentais" (European Parliament, 2020, tradução nossa). O autor desse relatório e membro do Parlamento Europeu, Ibán García del Blanco, ainda afirmou que "o cidadão está no centro dessa proposta" (European Parliament, 2020, tradução nossa). Nesse sentido, Axel Voss, também membro do Parlamento e autor do relatório sobre um regime de responsabilidade civil para a inteligência artificial, esclarece que a meta é garantir a proteção dos europeus e ao mesmo tempo oferecer às empresas a segurança jurídica necessária para promover a inovação.

Em abril de 2021, a Comissão Europeia propôs a primeira estrutura regulatória de IA para a União Europeia. Em dezembro de 2023, a Comissão acolheu o acordo político sobre a lei. O Ato de IA (AIA)[8] trata sobre os possíveis riscos decorrentes do uso de ferramentas de inteligência artificial, como riscos à saúde, à segurança e aos direitos fundamentais dos cidadãos. Assim, ela proporciona tanto aos desenvolvedores quanto aos distribuidores desses sistemas requisitos e obrigações em relação aos usos específicos da IA.

A ideia da lei de inteligência artificial surgiu no contexto da Conferência sobre o Futuro da Europa (COFE), que ocorreu entre abril de 2021 e maio de 2022. A conferência foi composta por diálogos e debates dos quais qualquer cidadão europeu podia participar. Assim, foi criado um espaço onde as pessoas puderam compartilhar ideias para ajudar a estabelecer as ações futuras para o continente. Ao fim, o evento contou com o envio de 49 propostas, sendo a lei de IA uma resposta à quatro propostas, especificamente[9].

---

[8] Para acessar e explorar o conteúdo do AIA de forma interativa, consulte o *The AI Act Explorer*, uma plataforma desenvolvida pela UE. Disponível em: <https://artificialintelligenceact.eu/ai-act-explorer/>. Acesso em: 2 nov. 2024.

[9] Proposta 12, medida 10: "identificar e desenvolver setores estratégicos, incluindo espaço, robótica e IA"; proposta 33, medida 5: "garantia da conformidade de empresas não europeias com as regras europeias de proteção de dados"; proposta 35, medida 3: "garantir a supervisão humana dos processos de tomada de decisão

Nesse contexto, no entanto, surgiu certo descontentamento entre grandes empresas. Em 2023, executivos de mais de 150 empresas assinaram uma carta aberta argumentando contra a regulamentação de IA desenvolvida na União Europeia[10]. Os signatários afirmaram que as leis impostas sobre sistemas de IA seriam excessivamente restritivas, podendo afetar negativamente a competitividade tecnológica da Europa e diminuindo as chances do bloco assumir um papel de liderança no âmbito da tecnologia. Como resposta à carta, membro do Parlamento Europeu e um dos atores mais importantes no desenvolvimento da política de IA europeia, Dragoș Tudorache lamentou que a pressão de algumas poucas empresas descontentes estivesse influenciando outras.

Ainda, Tudorache, defendendo o Ato de IA, afirmou que a legislação europeia proporciona "um processo liderado pelo setor para definir padrões, governança com o setor na mesa e um regime regulatório leve que exige transparência. Nada mais" (Weatherbed, 2023, tradução nossa). À época, a OpenAI, empresa responsável pelo ChatGPT, também protestou contra o regulamento. O CEO da empresa, Sam Altman, havia ameaçado deixar o mercado europeu caso não fosse possível aderir às normas propostas. Mais tarde, no entanto, esse posicionamento foi reavaliado e Altman afirmou que não deixaria a UE. Ademais, em 2024, a OpenAI prometeu cumprir o Ato de IA:

> Acreditamos que o foco principal do Pacto de IA na alfabetização, adoção e governança da IA tem como alvo as prioridades certas para garantir que os ganhos da IA sejam amplamente distribuídos. Além disso, eles estão alinhados com nossa missão de fornecer tecnologias seguras e de ponta que beneficiem a todos. [...] A OpenAI tem o compromisso de cumprir a Lei de IA da UE e trabalhará em estreita colaboração com o novo Escritório de IA da UE à medida que a lei for implementada. Nos próximos meses, continuaremos a preparar a documentação técnica e outras orientações para provedores e implantadores finais de nossos modelos GPAI [IA de uso geral], ao mesmo tempo em que promovemos a segurança e a proteção dos modelos que fornecemos no mercado europeu e fora dele (OpenAI, 2024, tradução nossa).

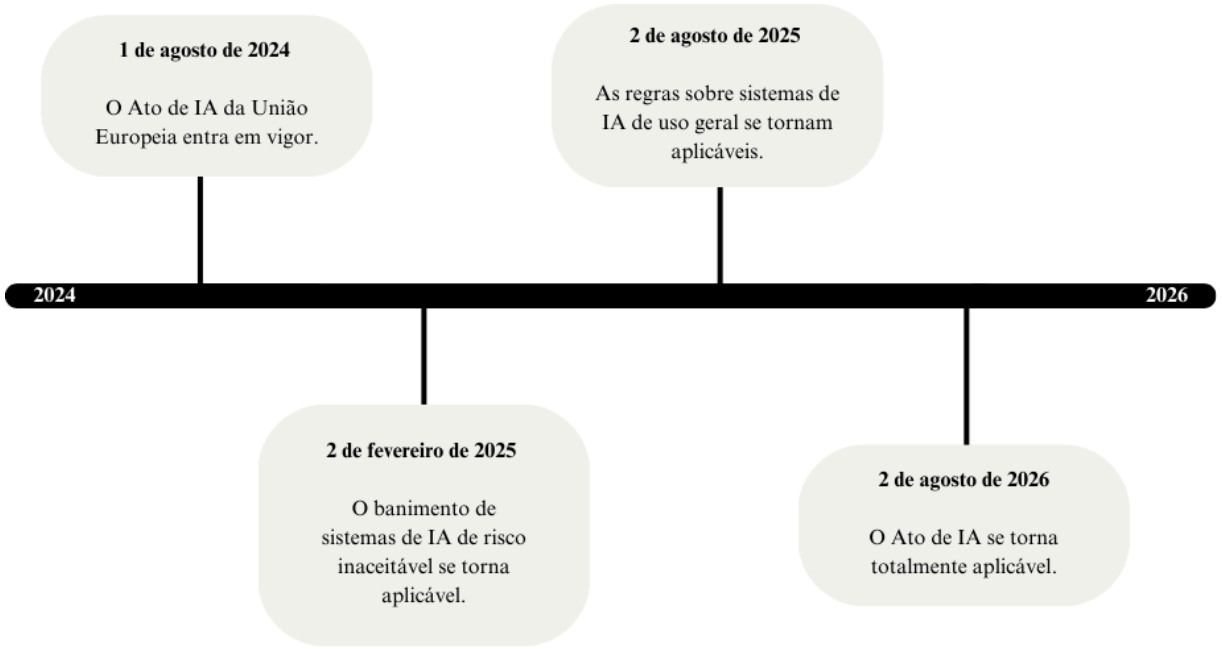
---

que envolvam inteligência artificial no local de trabalho e a transparência dos algoritmos usados; considerar os impactos negativos da vigilância digital limitada no local de trabalho; informar e consultar os trabalhadores antes da introdução de tecnologias digitais que afetem as condições de trabalho; garantir que novas formas de trabalho, como o trabalho em plataforma, respeitem os direitos dos trabalhadores e ofereçam condições de trabalho adequadas"; e medida 8: "Utilizar plenamente o potencial do uso confiável e responsável da inteligência artificial, usar o potencial da tecnologia blockchain e dos serviços em nuvem, definindo salvaguardas e padrões que garantam transparência, interoperabilidade, gerem confiança, aumentem a facilidade de uso e evitem algoritmos discriminatórios ou tendenciosos"; e proposta 37, medida 3: "Fazer maior uso da inteligência artificial e das tecnologias de tradução para contornar as barreiras linguísticas, garantindo a acessibilidade e a usabilidade de todas as ferramentas digitais para pessoas com deficiência" (European Parliament, 2024a, p. 55-80, tradução nossa).

[10] *Open letter to the representatives of the European Commission, the European Council and the European Parliament*. Disponível em: <https://www.igizmo.it/wp-content/uploads/2023/06/Open-Letter-EU-AI-Act-and-Signatories.pdf?trk=public_post_comment-text>. Acesso em: 26 out. 2024.

Em março de 2024, o Parlamento Europeu adotou o Ato de IA, que entrou em vigor em agosto do mesmo ano, mas terá um prazo de até 24 meses até se tornar totalmente aplicável. No entanto, segundo a Comissão Europeia, sistemas de IA cujo risco é considerado inaceitável já serão afetados pela lei após 6 meses, enquanto as regras sobre sistemas de IA de uso geral – modelos de IA capazes de executar tarefas, como geração de texto, similares aos humanos – serão aplicadas após 12 meses (European Commission, 2024a). A Figura 1 mostra uma linha do tempo do Ato de IA.

Figura 1: Linha do tempo do Ato de IA



Fonte: Elaboração própria com base em European Commission (2024a).

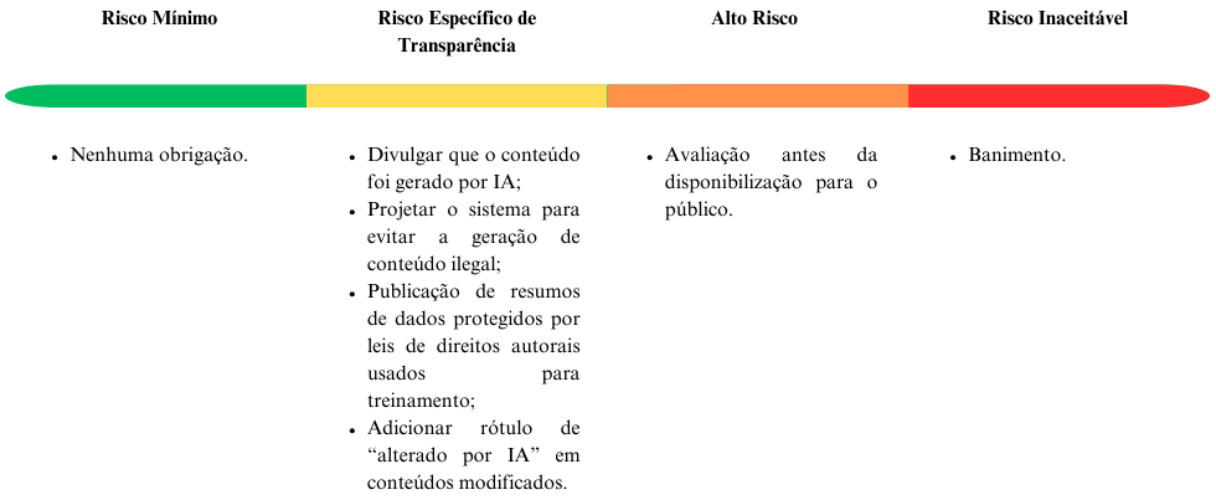
O ato se trata da primeira lei regulatória rígida propriamente implementada, marcando um importante momento no desenvolvimento da governança global de inteligência artificial. Nesse contexto, os Estados-membros da União Europeia têm até agosto de 2025 para designar autoridades nacionais competentes para supervisionar a implementação da lei, além de fiscalizar o mercado para garantir que esteja de acordo com a legislação.

3.2. ESTRUTURA E OBJETIVOS DO ATO DE IA

A prioridade do Parlamento é garantir que os sistemas de IA utilizados na União Europeia sejam “seguros, transparentes, rastreáveis, não discriminatórios e ecológicos”

(European Parliament, 2024b, tradução nossa). Sob essas regras, sistemas de inteligência artificial deveriam passar por uma análise de riscos para que fossem disponibilizados para o público. Assim, esses sistemas podem ser classificados como possuindo um "risco mínimo", "risco específico de transparência", "alto risco" ou até mesmo "risco inaceitável", como indicado na Figura 2.

Figura 2: Gradação de risco do Ato de IA



Fonte: Elaboração própria com base em European Commission (2024a).

Os sistemas classificados como "risco mínimo", como filtros de spam em e-mails e jogos de videogame que utilizam IA não terão nenhuma obrigação para com o Ato de IA. No entanto, empresas podem adotar códigos de conduta adicionais da legislação voluntariamente, se assim desejarem (European Commission, 2024a).

Por outro lado, aos sistemas classificados como possuindo um "risco específico de transparência", por exemplo no caso de sistemas de IA generativa, como o ChatGPT, foi imposta a obrigatoriedade de transparência do sistema, incluindo: divulgar que o conteúdo foi gerado por IA; projetar o sistema para evitar a geração de conteúdo ilegal; publicação de resumos de dados protegidos por leis de direitos autorais usados para treinamento. Ademais, qualquer conteúdo que seja modificado por IA, como imagens, áudios e vídeos, deve deixar explícito que houve alterações, adicionando um rótulo de "alterado por IA" (European Commission, 2024a).

Já os sistemas classificados como "alto risco" devem ser avaliados antes de serem disponibilizados para o público. Seriam incluídos nessa categoria: sistemas de IA utilizados em produtos abrangidos pela legislação de segurança dos produtos da União Europeia;

sistemas de IA que se enquadram em áreas específicas que terão de ser registadas numa base de dados da União Europeia, como: gestão e operação de infraestrutura crítica; educação e formação profissional; emprego, gestão de trabalhadores e acesso ao trabalho independente; acesso e usufruto de serviços privados essenciais e serviços e benefícios públicos; aplicação da lei; gestão da migração, do asilo e do controle das fronteiras; assistência na interpretação jurídica e aplicação da lei (European Commission, 2024a).

Por fim, os sistemas classificados como "risco inaceitável" devem ser banidos. Nessa categoria, seriam incluídos: manipulação cognitivo-comportamental de pessoas ou grupos vulneráveis específicos; pontuação social: classificar as pessoas com base no comportamento, status socioeconômico ou características pessoais; identificação biométrica e categorização de pessoas; sistemas de identificação biométrica remota e em tempo real, como reconhecimento facial. De acordo com a proposta, podem haver exceções por razões de imposição da lei, a depender da gravidade da situação, como por exemplo o uso de sistemas de identificação biométrica remota e em tempo real (European Commission, 2024a).

A Comissão afirma, ainda, que as empresas que não cumprirem as normas do regulamento serão multadas. A aprovação da lei causou entusiasmo entre lideranças europeias. O ministro belga da digitalização, Mathieu Michel, afirmou em um comunicado que

> esta lei histórica, a primeira do gênero no mundo, aborda um desafio tecnológico global que também cria oportunidades para as nossas sociedades e economias. Com a Lei da IA, a Europa enfatiza a importância da confiança, da transparência e da responsabilização ao lidar com novas tecnologias, ao mesmo tempo que garante que esta tecnologia em rápida mudança possa florescer e impulsionar a inovação europeia (European Council, 2024, tradução nossa).

Com uma estrutura regulatória sólida, baseada na transparência e nos direitos humanos, a União Europeia busca criar um ambiente benéfico de IA, melhorando a qualidade de vida dos cidadãos em áreas como saúde, transporte e outros serviços públicos. Além disso, também podem decorrer benefícios de maior produtividade de fabricação para empresas e maior eficiência em áreas como energia, impulsionando o desenvolvimento sustentável.

Além disso, em maio de 2024, foi instituído o Escritório de IA da Comissão Europeia, órgão responsável pela implementação da lei em nível da UE, além de aplicar as regras a sistemas de IA de uso geral. Segundo o Ato de IA, o escritório será apoiado pelos países membros da UE, que o auxiliarão e falicilitarão o cumprimento das suas funções. Ainda, por meio do Escritório, a Comissão Europeia patrocinará o desenvolvimento de habilidades e competências no campo da IA (European Parliament, 2024c).

Em complemento ao Escritório de IA, outros três órgãos de caráter consultivo irão apoiar a implementação das normas. O Conselho Europeu de Inteligência Artificial, presidido por um representante de um dos Estados-membros, será responsável por garantir a aplicação uniforme do Ato de IA em todas as nações da UE e também atuará como o principal órgão de cooperação entre a Comissão e os países do bloco. Cada um dos Estados-membros terá um representante na diretoria do Conselho, sendo que esses representantes serão encarregados de coordenar e regulamentar a IA em seus respectivos países. Ainda, o Conselho contará com a presença de observadores da Autoridade Europeia para a Proteção de Dados e do Escritório de IA (European Parliament, 2024c).

Além do Conselho, será implementado um painel científico de especialistas independentes. Esse painel será composto por especialistas com conhecimentos avançados em IA e independentes de fornecedores de sistemas de IA. O painel será responsável por fornecer consultoria técnica e informações sobre a aplicação da lei, além de emitir alertas para o Escritório de IA sobre os riscos associados aos modelos de IA de uso geral. Segundo o Ato de IA, os especialistas serão imparciais e confidenciais, além de não receberem instruções de nenhum terceiro. Além disso, eles não precisarão manifestar qualquer interesse, para evitar conflitos (European Parliament, 2024c).

E, por fim, um fórum consultivo, composto por um conjunto diversificado de partes interessadas, como startups, pequenas e médias empresas, a sociedade civil e o setor acadêmico. O fórum será responsável por orientar o Conselho e a Comissão, fornecendo conhecimento técnico e consultoria. Algumas organizações serão membros permanentes, mas a Comissão indicará também membros temporários que sejam experientes em IA. Segundo o Ato de IA, o fórum elegerá dois co-presidentes, criará suas próprias regras e se reunirá pelo menos duas vezes por ano. Por fim, será disponibilizado anualmente um relatório sobre suas atividades, que estará disponível publicamente (European Parliament, 2024c).

Em relação ao tema de *deep fakes* especificamente, o Ato de IA oferece uma definição para o termo. Segundo o artigo 3, parágrafo 60º, "*deep fake*" significa "conteúdo de imagem, áudio ou vídeo gerado ou manipulado por IA que se assemelha a pessoas, objetos, lugares, entidades ou eventos existentes e que, para uma pessoa, pareceria falsamente autêntico ou verdadeiro" (European Parliament, 2024c, tradução nossa). Ainda, no artigo 50, parágrafo 4º, o Ato de IA obriga os aplicadores de sistemas de IA que produzam *deep fakes* a deixar explícito que o conteúdo foi gerado ou manipulado artificialmente. No entanto, essa obrigação não se aplica em casos de uso autorizado para fins legais. Por fim, o artigo determina que *deep fakes* utilizadas com o propósito de entretenimento – como em programas ou outras

obras nitidamente culturais – possuem obrigações de transparência. Assim, nesses casos, deve ser divulgada a existência de conteúdo gerado ou manipulado, de forma que não prejudique a experiência do público ao assistir a obra.

No entanto, Wachter (2024) critica as limitações e as brechas deixadas pelo Ato de IA, afirmando que as obrigações deixaram muito a desejar. No caso das obrigações relativas a um risco de transparência, a autora comenta que sistemas de IA nessa classificação, como chatbots, já causaram grandes danos no passado, citando um exemplo no qual o sistema aconselhou pessoas a tirarem suas próprias vidas. Portanto, afirma a autora, a transparência não é suficiente para resolver esses problemas. Já quanto aos sistemas de alto risco, a autora considera que a lei omite áreas relevantes como "IA usada na mídia, sistemas de recomendação, ciência e academia, a maior parte das finanças e do comércio, a maioria dos tipos de seguro e aplicativos específicos voltados para o consumidor, como chatbots e algoritmos de preços" (Wachter, 2024, p. 682, tradução nossa), o que apresentam risco considerável.

Por fim, a autora aponta que muitos sistemas que poderiam ter sido considerados de risco inaceitável não foram contemplados pela lei, como identificação biométrica remota e reconhecimento de emoções, mesmo se tratando de sistemas pouco precisos e que muitas vezes causam impactos negativos ao direito de liberdade de expressão, privacidade, protesto e reunião. Além disso, a autora realça como não foi proibido que empresas europeias vendessem produtos de IA proibidos para outros países, o que ela entende como sendo uma grande oportunidade perdida.

3.3. IMPACTO DO ATO DE IA NAS RELAÇÕES INTERNACIONAIS

A União Europeia, com o Ato de IA, busca se tornar um líder global no segmento da inteligência artificial segura. O pioneirismo na implementação de uma lei de regulamentação de IA coloca o bloco em uma posição única. É esperado que a legislação influencie outras iniciativas ao redor do mundo, tendência já observada em outros momentos nos quais a UE serviu de exemplo para demais países e corporações, fenômeno que recebeu o nome de "Efeito Bruxelas". Segundo Bradford (2019), o Efeito Bruxelas se refere ao "poder unilateral da UE de regular os mercados globais" (Bradford, 2019, p. xiv, tradução nossa). Apesar dos desafios que o bloco vem passando, Bradford (2019) argumenta que a União Europeia continua sendo uma potência influente e que provavelmente o será durante muito tempo. A autora demonstra como as regulamentações europeias em diversos setores continuam

influenciando produtos no mundo todo. A razão desse fenômeno, afirma a autora, não é apenas o tamanho do mercado europeu, mas a arquitetura institucional construída no continente, com jurisdições rigorosas e muito respeitadas pelo resto do mundo.

Dessa maneira, mesmo com as dificuldades enfrentadas pela UE, ainda é válido esperar a manifestação do Efeito Bruxelas. Siegmann e Anderljung (2022) afirmam que um efeito Bruxelas é provável com o Ato de IA e que, caso esse efeito seja significativo, pode haver um esforço global por uma regulamentação mais rígida para a IA. Nesse contexto, os autores também diferenciam dois conceitos: o Efeito Bruxelas *de facto* e o Efeito Bruxelas *de jure*.

Um Efeito Bruxelas *de facto* ocorre quando empresas cumprem as normas da UE voluntariamente mesmo fora das fronteiras do bloco. Siegmann e Anderljung (2022) argumentam que isso é esperado, uma vez que, para muitas empresas multinacionais, é mais lucrativo oferecer o mesmo produto tanto dentro da UE quanto no resto do mundo, ao invés de oferecer duas versões distintas. Segundo os autores, isso é provável por conta da alta capacidade regulatória da UE, que têm maiores chances de produzir leis bem elaboradas e que não são excessivamente complicadas de se cumprir.

Já um Efeito Bruxelas *de jure* ocorre quando jurisdições estrangeiras adotam leis influenciadas pelas leis europeias. Siegmann e Anderljung (2022) justificam que isso também é provável, visto que jurisdições estrangeiras podem esperar que a regulamentação europeia seja de alta qualidade e compatível com seus próprios interesses. Ademais, a União Europeia pode incentivar que outros atores adotem regulamentações semelhantes através da participação em instituições e negociações comerciais. Assim, os autores acreditam que a difusão *de jure* ocorreria principalmente em grandes parceiros comerciais do bloco europeu, uma vez que a não-conformidade com a lei europeia traria atritos no comércio.

Por outro lado, em um ambiente de intensa competição global no setor de inteligência artificial, países como os Estados Unidos e a China estão trabalhando em suas próprias leis de regulamentação de IA. Sobre o caso chinês, Chen e Gao (2024) afirmam que a China vem desenvolvendo uma abordagem própria de regulamentação do ciberespaço, diferente das abordagens de países ocidentais. Os autores mostram que a China, insatisfeita com os problemas que identificou nas normas da governança cibernética global, busca impulsionar três valores fundamentais: soberania cibernética, multilateralismo e o equilíbrio entre segurança e desenvolvimento (Chen; Gao, 2024, p. 2439, tradução nossa).

Além disso, tratando dos mecanismos de difusão de normas, Chen e Gao (2024) revelam como a China utiliza ações de socialização em organizações internacionais e

estratégias de incentivo positivo para disseminar suas normas cibernéticas. Um exemplo de socialização apresentado pelos autores ocorre no âmbito dos BRICS, onde a convergência de ideias entre os países do bloco proporciona uma visão favorável à governança cibernética visada pela China. Essas ações de socialização, complementam os autores, geralmente são acompanhadas de mecanismos de incentivo positivo, que podem ser divididos em duas categorias: fornecimento de apoio financeiro e infraestrutura digital e facilitação da troca de informações e ações colaborativas.

Nesse contexto, surgiu a ideia de que o Efeito Bruxelas pode ser limitado, diminuindo a influência das decisões europeias no assunto. Mesmo assim, Siegmann e Anderljung (2022) deduzem que o Efeito Bruxelas provavelmente será mais significativo do que o "Efeito Washington" ou o "Efeito Pequim":

> É improvável que os EUA implementem uma legislação mais rigorosa do que a da UE, o que torna improvável um Efeito Washington. Pequim terá dificuldades para criar um Efeito Pequim de facto, pois as empresas geralmente já oferecem produtos especificamente para o mercado chinês, embora possa haver um Efeito Pequim de facto por meio de exportações de empresas chinesas. Há alguma chance de vermos um Efeito Pequim de jure em relação aos países que compartilham as metas regulatórias do Partido Comunista Chinês (Siegmann e Anderljung, 2022, p. 5, tradução nossa).

No entanto, os autores reconhecem que, caso a conformidade com a regulamentação europeia seja muito custosa para as empresas, ou a qualidade do produto de IA diminua – se funções forem limitadas propositalmente, reduzindo a funcionalidade –, existe a possibilidade de uma diminuição na demanda, levando a um encolhimento do mercado europeu e tornando as empresas mais dispostas a abandoná-lo. Assim, haveria também um incentivo para oferecer produtos que não estejam de acordo com a regulamentação europeia.

De qualquer modo, no contexto atual de regulamentação de IA, a responsabilidade recai sobre a União Europeia. Siegmann e Anderljung (2022) esperam que jurisdições de todo o mundo experimentem um Efeito Bruxelas *de facto*, mesmo que parcialmente. Assim, é particularmente importante que a regulamentação europeia esteja preparada para o futuro e eventuais transformações no campo da IA – como de fato foi pensado o Ato de IA –, para adaptar-se conforme necessário. Por fim, os autores defendem que a comunidade global da IA deve investir em pesquisa em tópicos como explicabilidade, justiça, transparência, robustez e supervisão humana, para que seja possível auxiliar na orientação dos esforços regulatórios da UE (Siegmann e Anderljung, 2022, p. 5, tradução nossa). Dessa forma, o Ato de IA pode ser visto como um apelo para encorajar formuladores de políticas ao redor do mundo a incentivar o desenvolvimento de pesquisas no campo da IA.

CONSIDERAÇÕES FINAIS

Neste trabalho, procuramos responder qual é o papel da governança global na busca por mitigar os riscos do uso da inteligência artificial para desinformação. Para responder essa pergunta, refletimos sobre a IA e os possíveis riscos de seu uso no contexto da disseminação de desinformação online, além de analisar as ações da comunidade global no sentido de regulamentação dessa tecnologia.

Assim sendo, buscamos primeiramente entender o significado do termo "inteligência artificial". Para alcançar esse objetivo, analisamos as diversas concepções e definições do termo, além de entender como a IA se desenvolveu ao longo do tempo. Dessa maneira, pudemos notar que ferramentas de inteligência artificial, por mais que não fossem tão perceptíveis, já fazem parte do nosso cotidiano há algum tempo, como em filtros de spam em e-mails, navegadores de pesquisa na internet, etc. No entanto, identificamos como as discussões que concernem à IA vêm aumentando nos últimos anos, em um verdadeiro hype da IA, principalmente com a recente democratização desses sistemas, proporcionada pelo fácil acesso a chatbots e ferramentas de geração de imagem e vídeo. Nessa perspectiva, percebemos como o risco do uso mal-intencionado dessa tecnologia aumentou, especialmente por conta dos *deep fakes*, que podem trazer consequências negativas para as relações internacionais.

Logo, ressaltamos como a questão da regulamentação de sistemas de IA se tornou uma prioridade. Exploramos o tema da governança global, suas definições e como ele se relaciona com o campo das RI e da IA. Discutimos as perspectivas da comunidade acadêmica sobre o tema, apontando os principais obstáculos para uma governança efetiva, bem como algumas soluções possíveis para lidar com esses problemas. Em complemento, selecionamos e examinamos alguns dos principais atores internacionais que propõem iniciativas de regulamentação de IA, como é o caso de organizações internacionais e blocos regionais. Diante disso, pudemos notar que a maior parte das iniciativas propostas até o momento se concentram apenas em fornecer recomendações, com pouco avanços no estabelecimento de normas sólidas para a IA.

Por fim, destacamos o caso da União Europeia, o ator responsável pelos avanços mais significativos para a regulamentação da IA. Examinamos o Ato de IA, as inspirações por trás dessa ideia, evidenciadas na Conferência sobre o Futuro da Europa (COFE), além de evidenciar o processo de formulação da lei e os descontentamentos de grandes empresas com as exigências do regulamento. Apresentamos os principais pontos do Ato de IA, como ele

classifica sistemas de IA baseado em riscos, mas também críticas quanto às suas limitações e brechas. Ainda, discutimos sobre as possíveis implicações da adoção do Ato de IA para o resto do mundo.

Como discutido pela literatura, é esperado que os desdobramentos ocorridos na União Europeia tenham influência sobre outras regiões – o chamado "Efeito Bruxelas" –, estabelecendo padrões a serem seguidos por boa parte da comunidade internacional. Conforme o apresentado, é provável que o Efeito Bruxelas se manifeste tanto *de facto*, pois pode ser mais lucrativo vender um produto padronizado ao invés de versões distintas, quanto *de jure*, por conta da expectativa de alta qualidade da lei europeia e dos interesses comuns de seus parceiros comerciais.

Apesar da crise vivenciada pela UE, a expectativa é de que o bloco continue em posição de relevância no sistema internacional pelos próximos anos. Assim, notamos que o Efeito Bruxelas é mais provável do que o Efeito Washington ou o Efeito Pequim. Logo, é esperado que o Ato de IA exerça um papel crucial no âmbito da governança global de IA, se tornando um modelo de legislação e um incentivo para outras regiões do mundo a adotarem medidas de regulamentação de IA. Portanto, é importante observar com atenção as repercussões do Ato de IA, como ele se mantém diante dos novos desafios e quais as lições que podem ser aprendidas com essa abordagem.

Diante do exposto, entendemos que a governança global tem papel fundamental na mitigação dos riscos do uso de IA para desinformação. O atual modelo de governança é insuficiente para lidar com os novos desafios proporcionados pelo desenvolvimento de tecnologias de IA. Os atores da governança global são cada vez mais heterogêneos e as instituições são numerosas, intensificando uma fragmentação institucional, causando dificuldades de coesão e reduzindo a eficácia de regimes. No entanto, a comunidade internacional tem demonstrado profundo interesse em combater os males da desinformação causada por IA, o que é evidenciado pelas diversas iniciativas que, mesmo em sua maioria atuando apenas com recomendações, comprovam as disposições por uma governança global dessa tecnologia.

Portanto, concluímos que é necessário que os esforços no sentido de uma governança global de IA sejam intensificados. Uma governança efetiva, que seja inclusiva e que respeite os direitos humanos, é crucial para solucionar a potencial crise trazida pela democratização da inteligência artificial. Nesse sentido, muito pode ser aproveitado do regionalismo para lidar com essa crise, tomando como modelo a União Europeia e seus esforços por uma legislação mais sólida de IA. A comunidade internacional tem, neste momento, a chance de adotar

medidas de regulamentação inspiradas no Ato de IA, o que seria um grande passo para a governança de IA.

Ademais, é visível o efeito negativo que o uso mal-intencionado de *deep fakes* traria para a sociedade, caso as ferramentas que possibilitam sua criação não sejam propriamente regulamentadas. Nas mãos de atores maliciosos, a desinformação causada por *deep fakes* pode se tornar uma ameaça a processos democráticos, eleições e segurança nacional. Logo, reconhecemos o valor do progresso realizado até o momento no âmbito da governança global de IA, mas é essencial que as pesquisas sobre o tema da inteligência artificial e desinformação continuem, e que os esforços de regulamentação se desenvolvam para acompanhar as transformações tecnológicas.

REFERÊNCIAS

AFRICAN UNION. **African Ministers Adopt Landmark Continental Artificial Intelligence Strategy, African Digital Compact to drive Africa's Development and Inclusive Growth**. Addis Ababa, 2024. Disponível em: <https://au.int/en/pressreleases/20240617/african-ministers-adopt-landmark-continental-artificial-intelligence-strategy>. Acesso em: 15 set. 2024.

ANYOHA, R. **The History of Artificial Intelligence**. Harvard University, The School of Arts and Sciences, 2017. Disponível em: <https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/>. Acesso em: 1 set. 2024.

ASARO, P. Algorithms of Violence: Critical Social Perspectives on Autonomous Weapons. **Social Research: An International Quarterly**, v. 86, n. 2, p. 537–555, jun. 2019.

ASEAN Secretariat. **ASEAN Guide on AI Governance and Ethics**. Jacarta: ASEAN Secretariat, 2024. Disponível em: <https://asean.org/book/asean-guide-on-ai-governance-and-ethics/>. Acesso em: 28 jun. 2024.

ASSOCIATION OF SOUTHEAST ASIAN NATIONS. **ASEAN initiates regional discussion on generative AI Policy**. ASEAN Main Portal, 2023. Disponível em: <https://asean.org/asean-initiates-regional-discussion-on-generative-ai-policy/>. Acesso em: 28 jun. 2024.

BARNETT, Michael N.; PEVEHOUSE, Jon C. W.; RAUSTIALA, Kal. Introduction: The Modes of Global Governance. In: BARNETT, Michael N.; PEVEHOUSE, Jon C. W.; RAUSTIALA, Kal (Orgs.). **Global Governance in a World of Change**. Cambridge: Cambridge University Press, 2021. Disponível em: <https://www.cambridge.org/core/books/global-governance-in-a-world-of-change/global-governance-in-a-world-of-change/0157187AA4C3F7A775D591BB71013089>. Acesso em: 28 out. 2024.

BELLMAN, R. E. **An Introduction to Artificial Intelligence**: Can Computers Think? San Francisco: Boyd & Fraser Publishing Company, 1978.

BODE, Ingvild. AI Technologies and International Relations: Do We Need New Analytical Frameworks? **The RUSI Journal**, v. 169, n. 5, 2024. Disponível em: <https://doi.org/10.1080/03071847.2024.2392394>. Acesso em: 11 set. 2024.

BORRÁS, Susana; EDLER, Jakob. **The Governance of Socio-Technical Systems: Explaining Change**. Cheltenham, Northampton: Edward Elgar Publishing, 2014.

BOUSQUET, A. The Persistent Appeal of Chaoplexic Warfare: Towards an Autonomous S(war)m Machine? In: GRUSZCZAK, A.; KAEMPF, S. (Orgs.). **The Routledge Handbook of the Future of Warfare**. London and New York: Routledge, 2023.

BRADFORD, Anu. **The Brussels Effect: How the European Union Rules the World**. New York: Oxford University Press, 2019. Disponível em: <https://academic.oup.com/book/36491>. Acesso em: 27 out. 2024.

BRASIL. Presidente (2023-: Luiz Inácio Lula da Silva). **Discurso na abertura da 79ª Assembleia Geral da ONU**. Nova York, 24 set. 2024a. Disponível em: <https://www.gov.br/planalto/pt-br/acompanhe-o-planalto/discursos-e-pronunciamentos/2024/09/discurso-do-presidente-lula-na-abertura-da-79a-assembleia-geral-da-onu-em-nova-york>. Acesso em: 2 nov. 2024.

BRASIL. Presidente Lula recebe chanceler russo Sergey Lavrov. **Portal do Planalto**, Brasília, 24 fev. 2024b. Disponível em: <https://www.gov.br/planalto/pt-br/acompanhe-o-planalto/noticias/2024/02/presidente-lula-recebe-chanceler-russo-sergey-lavrov>. Acesso em: 28 jun. 2024.

BUDRYK, Zack. **Fake videos edited to make Pelosi appear drunk spread on social media**. The Hill, 2019. Disponível em: <https://thehill.com/policy/cybersecurity/445311-fake-videos-spread-on-social-%20media-edited-to-make-pelosi-appear-drunk./>. Acesso em: 1 set. 2024.

BÜTHE, T., DJEFFAL, C., LÜTGE, C., MAASEN, S., & INGERSLEBEN-SEIP, N. von. Governing AI – attempting to herd cats? Introduction to the special issue on the Governance of Artificial Intelligence. **Journal of European Public Policy**, v. 29, n. 11, 1721–1752, 2022. Disponível em: <https://doi.org/10.1080/13501763.2022.2126515>. Acesso em: 15 set. 2024.

COMITÊ GESTOR DA INTERNET NO BRASIL. **Internet, democracia e eleições : guia prático para gestores públicos e usuários** / Núcleo de Informação e Coordenação do Ponto BR. São Paulo: Comitê Gestor da Internet no Brasil, 2018. Disponível em: <https://bibliotecadigital.acervo.nic.br/items/edfdfb30-e271-4cc2-9112-2e7b73a1756c>. Acesso em: 1 set. 2024.

CHAMAYOU, Gregoire. **Teoria do Drone**. São Paulo: Cosac Naify, 2015.

CHARNIAK, E.; MCDERMOTT, D. **Introduction to Artificial Intelligence**. Reading, Menlo Park, New York: Addison-Wesley, 1985.

CHEN, Xuechen; GAO, Xinchuchu. Norm diffusion in cyber governance: China as an emerging norm entrepreneur? **International Affairs**, v. 100, n. 6, p. 2419–2440, 2024. Disponível em: <https://doi.org/10.1093/ia/iiae237>. Acesso em: 11 nov. 2024.

CHESNEY, Robert; CITRON, Danielle K. **Disinformation on Steroids: The Threat of Deep Fakes**. Council on Foreign Relations, 2018. Disponível em: <https://www.cfr.org/report/deep-fake-disinformation-steroids>. Acesso em: 1 set. 2024.

CHHOTRAY, Vasudha; STOKER, Gerry. **Governance theory and practice: a cross-disciplinary approach**. Londres: Palgrave Macmillan, 2009.

DREZNER, Daniel W. Technological change and international relations. **International Relations**, v. 33, n. 2 2019. Disponível em: <https://journals.sagepub.com/doi/10.1177/0047117819834629>. Acesso em: 15 set. 2024.

EUROPEAN COMMISSION. **AI Act enters into force**. 2024. Disponível em: <https://commission.europa.eu/news/ai-act-enters-force-2024-08-01_en>. Acesso em: 11 out. 2024.

EUROPEAN COMMISSION. **COMMUNICATION FROM THE COMMISSION: Artificial intelligence for Europe**. EUR-Lex, 2018. Disponível em: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN>. Acesso em: 1 set. 2024.

EUROPEAN COMMISSION. **Conference on the Future of Europe**. Disponível em: <https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/new-push-european-democracy/conference-future-europe_en>. Acesso em: 28 jun. 2024.

EUROPEAN COMMISSION. **European Artificial Intelligence Act comes into force**. 2024a. Disponível em: <https://ec.europa.eu/commission/presscorner/detail/en/ip_24_4123>. Acesso em: 11 out. 2024.

EUROPEAN COUNCIL. **Artificial intelligence (AI) act: Council gives final green light to the first worldwide rules on AI**. Council of the EU, 2024. Disponível em: <https://www.consilium.europa.eu/en/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/>. Acesso em: 26 out. 2024.

EUROPEAN ECONOMIC AND SOCIAL COMMITTEE. **Opinion of the European Economic and Social Committee on 'Artificial intelligence — The consequences of artificial intelligence on the (digital) single market, production, consumption, employment and society'**. 2017. Disponível em: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016IE5369>. Acesso em: 13 nov. 2024.

EUROPEAN PARLIAMENT. **AI rules: what the European Parliament wants**. 2020. Disponível em: <https://www.europarl.europa.eu/topics/en/article/20201015STO89417/ai-rules-what-the-european-parliament-wants>. Acesso em: 18 out. 2024.

EUROPEAN PARLIAMENT. **Artificial Intelligence Act: MEPs adopt landmark law**. 2024a. Disponível em: <https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law>. Acesso em: 28 jun. 2024.

EUROPEAN PARLIAMENT. **Conference on the Future of Europe: REPORT ON THE FINAL OUTCOME**. 2022. Disponível em: <https://conference-followup.europarl.europa.eu/cmsdata/267078/Report_EN.pdf>. Acesso em: 28 jun. 2024.

EUROPEAN PARLIAMENT. **EU AI Act: first regulation on artificial intelligence**. 2024b. Disponível em: <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>. Acesso em: 28 jun. 2024.

EUROPEAN PARLIAMENT. **Texts adopted - Artificial Intelligence Act**. 2024c. Disponível em: <https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html#title2>. Acesso em: 28 jun. 2024.

EXECUTIVE OFFICE OF THE PRESIDENT. **The national artificial intelligence research and development strategic plan**. Washington, DC: National Science and Technology Council, 2016. Disponível em: <https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf>. Acesso em: 1 set. 2024.

GARCIA, Denise. Lethal Artificial Intelligence and Change: The Future of International Peace and Security. **International Studies Review**, v. 20, n. 2, p. 334-34, 2018.

GERRITS, André. Disinformation in International Relations: How Important Is It? **Security and Human Rights**, v. 29, n. 1-4, p. 3–23, 2018. Disponível em: <https://brill.com/view/journals/shrs/29/1-4/article-p3_3.xml?ebody=pdf-117260>. Acesso em: 11 set. 2024.

GILLESPIE, Tarleton. **A relevância dos algoritmos**. Revista Parágrafo. São Paulo, Brasil, v. 6, n. 1, p. 95-121, jan./abr. 2018. Disponível em: <https://revistaseletronicas.fiamfaam.br/index.php/recicofi/article/view/722/563>. Do original: The relevance of algorithms in Media Technologies: Essays on Communication, Materiality, and Society (MIT Press, 2014). Tradução: Amanda Jurno. Revisão: Carlos d'Andréa. Acesso em 1 set. 2024.

GLOBAL PARTNERSHIP ON ARTIFICIAL INTELLIGENCE. **About GPAI**. Disponível em: <https://gpai.ai/about/>. Acesso em: 12 jun. 2024.

GLOBAL PARTNERSHIP ON ARTIFICIAL INTELLIGENCE. **Social Media Governance Project: Summary of Work in 2023**. 2023. Disponível em: <https://gpai.ai/projects/responsible-ai/RAI03%20-%20Social%20Media%20Governance%20Project%20-%20Summary%20of%20Work%20in%202023.pdf>. Acesso em: 12 jun. 2024.

HARWELL, Drew. **Top AI researchers race to detect "deepfake" videos: "We are outgunned"**. Washington Post, 2019. Disponível em: <https://www.washingtonpost.com/technology/2019/06/12/top-ai-researchers-race-detect-deepfake-videos-we-are-outgunned/>. Acesso em: 1 set. 2024.

HAUGELAND, J. (Org.). **Artificial Intelligence**: The Very Idea. Cambridge: MIT Press, 1989.

HEYNS, Christof. Autonomous Weapons Systems: Living a dignified life and dying a dignified death. In: BHUTA, N., BECK, S. GEIB, R., LIU, HY., KREB, C. (Orgs.). **Autonomous Weapons Systems**: Law, Ethics, Policy, Cambridge: Cambridge University Press, 2016.

HO, Lewis; BARNHART, Joslyn; TRAGER, Robert; et al. **International Institutions for Advanced AI**, 2023. Disponível em: <https://arxiv.org/pdf/2307.04699>. Acesso em:: 6 set. 2024.

INTERNATIONAL TELECOMMUNICATION UNION. **AI for Good**, 2024a. Disponível em: <https://aiforgood.itu.int/>. Acesso em: 12 jun. 2024.

KNOTT, Alistair; PEDRESCHI, Dino; CHATILA, Raja; et al. Generative AI models should include detection mechanisms as a condition for public release. **Ethics and Information Technology**, v. 25, n. 4, p. 1–7, 2023. Disponível em: <https://doi.org/10.1007/s10676-023-09728-4>. Acesso em: 12 jun. 2024.

KUHLMANN, Stefan; STEGMAIER, Peter; KONRAD, Kornelia. The tentative governance of emerging science and technology—A conceptual introduction. **Research Policy**, v. 48, n. 5, p. 1091–1097, 2019. Disponível em: <https://doi.org/10.1016/j.respol.2019.01.006>. Acesso em: 17 jul. 2024.

KURZWEIL, R. **The Age of Intelligent Machines**. Cambridge: MIT Press, 1990.

MERCOSUL. **DECLARAÇÃO ESPECIAL DOS PRESIDENTES DO MERCOSUL SOBRE DEMOCRACIA E INTEGRIDADE DA INFORMAÇÃO EM AMBIENTES DIGITAIS**. 2023. Disponível em: <https://www.mercosur.int/pt-br/declaracao-especial-dos-presidentes-do-mercosul-sobre-democracia-e-integridade-da-informacao-em-ambientes-digitais/>. Acesso em: 28 jun. 2024.

MERCOSUL. **Fórum Internacional "Crimes virtuais e desinformação política: ameaças à integridade eleitoral"**. 2024. Disponível em: <https://www.mercosur.int/pt-br/forum-internacional-crimes-virtuais-e-desinformacao-politica-ameacas-a-integridade-eleitoral/>. Acesso em: 28 jun. 2024.

METZ, Rachel. **Deepfakes are now trying to change the course of war**. CNN, 2022. Disponível em: <https://edition.cnn.com/2022/03/25/tech/deepfakes-disinformation-war/index.html>. Acesso em: 1 set. 2024.

MILNER, Helen V.; SOLSTAD, Sondre Ulvund. Technological Change and the International System. **World Politics**, v. 73, n. 3. 2021. Disponível em: <10.1017/s0043887121000010>. Acesso em: 15 set. 2024.

MINISTÉRIO DOS DIREITOS HUMANOS E DA CIDADANIA. **Em plenária final, RAADH aprova Declaração de Princípios de Direitos Humanos no âmbito da Inteligência Artificial no Mercosul**. Gov.br, 2023. Disponível em: <https://www.gov.br/mdh/pt-br/assuntos/noticias/2023/novembro/em-plenaria-final-raadh-aprova-declaracao-de-principios-de-direitos-humanos-no-ambito-da-inteligencia-artificial-no-mercosul>. Acesso em: 28 jun. 2024.

MISHRA, Vibhu. **General Assembly adopts landmark resolution on artificial intelligence**. UN News, 2024. Disponível em: <https://news.un.org/en/story/2024/03/1147831#:~:text=The%20UN%20General%20Assembly%20on%20Thursday%20adopted%20a%20landmark%20resolution>. Acesso em: 22 set. 2024.

NILSSON, N. J. **Artificial Intelligence**: A New Synthesis. San Francisco: Morgan Kaufmann, 1998.

OPENAI. **A Primer on the EU AI Act: What It Means for AI Providers and Deployers**. OpenAI, 2024. Disponível em: <https://openai.com/global-affairs/a-primer-on-the-eu-ai-act/?form=MG0AV3>. Acesso em: 25 out. 2024.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **G7 Hiroshima Process on Generative Artificial Intelligence (AI): Towards a G7 Common Understanding on Generative AI**. Paris: OECD Publishing, 2023. Disponível em: <https://www.oecd-ilibrary.org/science-and-technology/g7-hiroshima-process-on-generative-artificial-intelligence-ai_bf3c0c60-en>. Acesso em: 12 jun. 2024.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449**. 2019. Disponível em: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>. Acesso em: 15 set. 2024.

PARTNERSHIP ON AI. **Partnership on AI**, 2024. About PAI. Disponível em: <https://partnershiponai.org/about/>. Acesso em: 12 jun. 2024.

PATERSON, Thomas; HANLEY, Lauren. **Political warfare in the digital age: cyber subversion, information operations and 'deep fakes'**. Australian Journal of International Affairs, 2020. Disponível em: <https://doi.org/10.1080/10357718.2020.1734772>. Acesso em: 1 set. 2024.

PERON, Alcides Eduardo dos Reis. **American way of war: o reordenamento sociotécnico dos conflitos contemporâneos e o uso de drones**. 2016. Tese de Doutorado. Tese de Doutorado em Políticas Científicas e Tecnológicas. Campinas: Universidade Estadual de Campinas.

POLYAKOVA, Alina; BOYER, Spencer. **THE FUTURE OF POLITICAL WARFARE: RUSSIA, THE WEST, AND THE COMING AGE OF GLOBAL DIGITAL COMPETITION**. Brookings – Robert Bosch Foundation, 2018. Disponível em: <https://www.brookings.edu/wp-content/uploads/2018/03/fp_20180316_future_political_warfare.pdf>. Acesso em: 1 set. 2024.

POOLE, D.; MACKWORTH, A. K.; GOEBEL, R. **Computational intelligence**: A logical approach. New York: Oxford University Press, 1998.

PRADO, Magaly. Inteligência artificial na cultura informativa e algoritmos de enganação. In: SANTAELLA, Lucia (Org.). **Inteligência Artificial & Redes Sociais**. São Paulo: Educ, 2019.

RICH, E.; KNIGHT, K. **Artificial Intelligence**. New York: McGraw-Hill, 1991.

ROBERTS, Huw; HINE, Emmie; TADDEO, Mariarosaria; FLORIDI, Luciano. Global AI governance: barriers and pathways forward. **International Affairs**, v. 100, n. 3, p. 1275–1286, 2024. Disponível em: <https://doi.org/10.1093/ia/iiae073>. Acesso em: 1 set. 2024.

ROTHKOPF, Joshua. **Deepfake Technology Enters the Documentary World**. The New York Times, 2020. Disponível em: <https://www.nytimes.com/2020/07/01/movies/deepfakes-documentary-welcome-to-chechnya.html>. Acesso em: 1 set. 2024.

RUSSELL, Stuart J.; NORVIG, Peter. **Artificial Intelligence: A Modern Approach**. Pearson Education, Inc., Upper Saddle River, New Jersey, 2009.

SHEARER, Elisa; GOTTFRIED, Jeffrey. **News Use Across Social Media Platforms 2017**. Pew Research Center, 2017. Disponível em: <https://www.pewresearch.org/journalism/2017/09/07/news-use-across-social-media-platforms-2017/>. Acesso em: 1 set. 2024.

SHIMABUKURO, Igor; MARQUES, Ana. **O que é deepfake? Conheça exemplos e entenda os riscos dessa tecnologia**. Tecnoblog, 2017. Em: <https://tecnoblog.net/264153/o-que-e-deep-fake-e-porque-voce-deveria-se-preocupar-com-isso/>. Acesso em: 29 out. 2024.

SIEGMANN, Charlotte; ANDERLJUNG, Markus. **The Brussels Effect and Artificial Intelligence: How EU regulation will impact the global AI market**. Centre for the GOVERNANCE OF AI, 2022. Disponível em: <https://arxiv.org/pdf/2208.12645>. Acesso em: 17 out. 2024.

SIMON, Herbert A. Artificial intelligence: an empirical science. **Artificial Intelligence**, v. 77, n. 1, p. 95–127, 1995. Disponível em: <https://doi.org/10.1016/0004-3702(95)00039-H>. Acesso em: 10 set. 2024.

SLOANE, Mona. Threading Innovation, Regulation, and the Mitigation of AI Harm: Examining Ethics in National AI Strategies. In: TINNIRELLO, Maurizio (Org.). **The Global Politics of Artificial Intelligence**. Boca Raton, Oxon: CRC Press, 2022. Disponível em: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4143375>. Acesso em: 1 set. 2024.

SONNE, Paul. **Fake Putin Speech Calling for Martial Law Aired in Russia**. The New York Times, 2023. Disponível em: <https://www.nytimes.com/2023/06/05/world/europe/putin-deep-fake-speech-hackers.html>. Acesso em: 1 set. 2024.

SPENCER, Michael. **Artificial intelligence hype is real**. Forbes Magazine, 2019. Disponível em: <https://www.forbes.com/sites/cognitiveworld/2019/02/25/artificial-intelligence-hype-is-real/>. Acesso em: 1 set. 2024.

STONE, Peter; BROOKS, Rodney; BRYNJOLFSSON, Erik; CALO, Ryan; ETZIONI, Oren; HAGER, Greg; HIRSCHBERG, Julia; KALYANAKRISHNAN, Shivaram; KAMAR, Ece; KRAUS, Sarit; Leyton-Brown, Kevin; PARKES, David; PRESS, William; SAXENIAN, AnnaLee; SHAH, Julie; TAMBE, Milind; TELLER, Astro. **Artificial Intelligence and Life in 2030**: One Hundred Year Study on Artificial Intelligence. Report of the 2015-2016 Study Panel. Stanford: Stanford University, 2016. Disponível em: <https://ai100.stanford.edu/2016-report>. Acesso em: 6 set. 2024.

SUWAJANAKORN, Supasorn; SEITZ, Steven M.; KEMELMACHER-SHLIZERMAN, Ira. Synthesizing Obama: Learning Lip Sync from Audio. **ACM Transactions on Graphics**, v. 36, n. 4, p. 1–13, 2017. Disponível em: <http://grail.cs.washington.edu/projects/AudioToObama/siggraph17_obama.pdf>. Acesso em: 1 set. 2024.

TAEIHAGH, Araz. Governance of artificial intelligence. **Policy and Society**, v. 40, n. 2, p. 137–157, 2021. Disponível em: <https://www.tandfonline.com/doi/epdf/10.1080/14494035.2021.1928377?needAccess=true>. Acesso em: 2 ago. 2024.

ULNICANE, Inga; KNIGHT, William; LEACH, Tonii; STAHL, Bernd; WANJIKU, Winter-Gladys. Governance of Artificial Intelligence: Emerging International Trends and Policy Frames. In: TINNIRELLO, Maurizio (Org.). **The Global Politics of Artificial Intelligence**. Boca Raton e Oxon: CRC Press, 2022. Disponível em: <https://www.taylorfrancis.com/books/edit/10.1201/9780429446726/global-politics-artificial-intelligence-maurizio-tinnirello>. Acesso em: 11 set. 2024.

ULNICANE, Inga; KNIGHT, William; LEACH, Tonii; STAHL, Bernd; WANJIKU, Winter-Gladys. Framing governance for a contested emerging technology: insights from AI policy. **Policy and Society**, v. 40, n. 2, p. 158–177, 2021. Disponível em: <https://doi.org/10.1080/14494035.2020.1855800>. Acesso em: 1 set. 2024.

VINNOVA. **Artificial Intelligence in Swedish Business and Society: Analysis of Development and Potential**. Vinnova - Sweden's Innovation Agency, 2018. Disponível em: <https://www.vinnova.se/contentassets/72ddc02d541141258d10d60a752677df/vr-18_12.pdf?cb=20180924082158>. Acesso em: 1 set. 2024.

WACHTER, Sandra. Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond. **Yale Journal of Law & Technology**, v. 26, n. 3, 2024. Disponível em: <https://yjolt.org/sites/default/files/wachter_26yalejltech671.pdf>. Acesso em: 27 out. 2024.

WAKEFIELD, Jane. **Guerra na Ucrânia: os 'presidentes deepfake 'usados na propaganda do conflito**. BBC News Brasil, 2022. Disponível em: <https://www.bbc.com/portuguese/internacional-60791955>. Acesso em: 1 set. 2024.

WARDLE, Claire; DERAKHSHAN, Hossein. **INFORMATION DISORDER: Toward an interdisciplinary framework for research and policy making**. Council of Europe, 2017. Disponível em: <https://rm.coe.int/information-disorder-toward-an-%20interdisciplinary-framework-for-researc/168076277c>. Acesso em: 1 set. 2024.

WEATHERBED, Jess. **European companies slam the EU's incoming AI regulations in open letter**. The Verge, 2023. Disponível em: <https://www.theverge.com/2023/6/30/23779611/eu-ai-act-open-letter-artificial-intelligence-regulation-renault-siemens>. Acesso em: 25 out. 2024.

WINSTON, P. H. **Artificial Intelligence**. Reading, Menlo Park, New York: Addison-Wesley, 1992.

APÊNDICE A: INSTITUIÇÕES INTERNACIONAIS E SUAS INICIATIVAS DE IA

Instituição	Iniciativa	Ano de criação	Tipos de membros	Enfoque
PAI	AI and Media Integrity Work	2016	Sociedade civil, empresas privadas e setor acadêmico[1]	Desenvolver ferramentas que auxiliem na divulgação de conhecimento acerca de IA, para criar um maior preparo da sociedade para lidar com questões referentes a essa tecnologia
ONU	AI for Good	2017	Governos, indústrias, agências da ONU, sociedade civil, organizações internacionais, setor acadêmico[2]	Sistemas de IA seguros, protegidos e confiáveis
OCDE	Recomendação sobre Inteligência Artificial	2019	Governos[3]	Inovação e confiança na IA, direitos humanos e valores democráticos, crescimento inclusivo, estado de direito, transparência, robustez, segurança e responsabilidade
GPAI	Responsible AI for social media governance	2020	Governos[4]	Preencher a lacuna entre a teoria e a prática em IA,

				apoiando investigação de ponta e atividades aplicadas em prioridades relacionadas com a IA
G7	Hiroshima AI Process	2023	Governos[5]	Desenvolvimento de IA segura e confiável
Mercosul	Declaração de Princípios de Direitos Humanos no âmbito da Inteligência Artificial no Mercosul	2023	Governos[6]	Desenvolvimento responsável da IA, igualdade e não discriminação, promoção da educação e formação digital, eliminação de segmentos étnico-raciais na IA, tecnologias de reconhecimento facial com transparência, privacidade e sem segmentos raciais
União Europeia	Ato de IA	2023	Governos[7]	Sistemas de IA seguros, transparentes, rastreáveis, não discriminatórios e ecológicos
ASEAN	ASEAN Guide on AI Governance and Ethics	2024	Governos[8]	Aumentar o conhecimento da população sobre os potenciais riscos e benefícios da inteligência artificial para

				que as pessoas possam se proteger de usos nocivos de sistemas de IA
União Africana	Estratégia Continental de Inteligência Artificial	2024	Governos[9]	Abordagem centrada nas pessoas, orientada para o desenvolvimento e inclusiva, garantia de salvaguardas adequadas e proteção contra ameaças
BRICS	—	—	Governos[10]	—

[1] Participação de 126 membros.
[2] Não há um número exato de participantes.
[3] A recomendação foi aderida por 36 membros da OCDE, 11 não-membros e pela UE.
[4] Participação de 28 países e da União Europeia.
[5] Participação dos 7 membros do grupo.
[6] Participação de representantes governamentais dos países membros e associados do Mercosul.
[7] Inclusão dos 27 Estados-membros.
[8] Participação dos 10 Estados-membros do bloco.
[9] Inclusão dos 55 Estados-membros.
[10] Países membros do BRICS.

Fonte: Elaboração própria com base em Partnership on AI (2024), International Telecommunication Union (2024a), Organisation for Economic Co-operation and Development (2019, 2023), Global Partnership on Artificial Intelligence (2020), Ministério dos Direitos Humanos e da Cidadania (2024), European Parliament (2024b), ASEAN Secretariat (2024), African Union (2024), Brasil (2024b).