DEPARTAMENTO DE DIREITO

MESTRADO EM DIREITO

ESPECIALIDADE EM CIÊNCIAS JURÍDICAS

UNIVERSIDADE AUTÓNOMA DE LISBOA

"LUÍS DE CAMÕES"

"HUMAN RIGHTS BY DESIGN" – OS DESAFIOS DA REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL

Dissertação para a obtenção do grau de Mestre em Direito

Autor: Gáudio Ribeiro de Paula

Orientador: Professor Doutor Pedro Gonçalo Tavares Trovão do Rosário

Número do candidato: 30001885

Janeiro de 2023

Lisboa

RESUMO

O caráter disruptivo da Inteligência Artificial (IA) já se evidenciou ao longo dos últimos anos. A contribuição da tecnologia para a melhoria das condições de vida das pessoas não pode afastar a preocupação com seus potenciais deletérios, que costumam se ocultar na densa névoa lançada pela ficção científica a antropomorfizar os agentes artificialmente inteligentes. Algumas de suas características quase intrínsecas tendem a dificultar a contenção de seus riscos atuais, como é o caso de sua opacidade, ubiquidade, complexidade e imprevisibilidade. Mesmo nos sistemas hodiernos, podem ser identificados relevantes impactos quanto aos direitos humanos fundamentais, tais com a vida, integridade física, dignidade da pessoa humana, isonomia liberdade de expressão e garantias laborais. Para conter tais vulnerabilidades, percebe-se um crescente esforço regulatório, em diversas frentes, com abordagens de "*hard law*", "*soft law*", autoregulação, co-regulação ou autoregulação regulada. Uma das mais promissoras estratégias consiste na incorporação de parâmetros estruturais alinhados aos direitos humanos na sua própria arquitetura algorítmica, de modo de obstar a violação dessas garantias mediante "travas" inseridas ao seu próprio "*design*" dos sistemas. Trata-se de uma concepção que alguns denominam "*Human rights by design*", para a implementação da qual cogitam-se de soluções metolócias como "*value specification*", "*coherent extrapolated volition*", controle de capacidades e seleção de motivações. De outro lado, emergem limitações de ordem epistemológica, além de desafios como a construção de consenso e a necessidade de supervisão. Da superação de tais dificuldades depende, em grande medida, a salvaguarda da dignidade humana em um cenário de possível ruptura na malha de proteção dos direitos fundamentais ante as implicações do célere avanço da IA em quase todas as atividades humanas.

PALAVRAS-CHAVE

Inteligência artificial – Riscos – Direitos Humanos – Regulação– Tecnoregulação

ABSTRACT

The disruptive nature of Artificial Intelligence (AI) has already been evident over the last few years. The contribution of technology to the improvement of people's living conditions cannot dispel concerns about its deleterious potential, which is usually hidden in the dense mist cast by science fiction that anthropomorphizes artificially intelligent agents. Some of its almost intrinsic characteristics tend to make it difficult to contain its current risks, such as its opacity, ubiquity, complexity and unpredictability. Even in today's systems, relevant impacts on fundamental human rights can be identified, such as life, physical integrity, human dignity, equality, freedom of expression and labor guarantees. To contain such vulnerabilities, a growing regulatory effort can be seen, on several fronts, with "hard law", "soft law", self-regulation, co-regulation or regulated self-regulation approaches. One of the most promising strategies is the incorporation of structural parameters aligned with human rights in its own algorithmic architecture, in order to prevent the violation of these guarantees through "locks" inserted into its own "design" of the systems. This is a conception that some call "Human rights by design", for the implementation of which methodological solutions are considered, such as "value specification", "coherent extrapolated volition", control of capacities and selection of motivations. On the other hand, limitations of an epistemological nature emerge, in addition to challenges such as building consensus and the need for supervision. The safeguarding of human dignity depends, to a large extent, on overcoming such difficulties in a scenario of possible disruption in the protection of fundamental rights as a result of the implications of the rapid advancement of AI in almost all human activities.

KEY WORDS

Artificial intelligence – Risks – Human Rights – Regulation – Technoregulation

SUMÁRIO

Lista de Tabelas

Lista de Figuras

INTRODUÇÃO

A formulação de modelos de regulação dos sistemas de inteligência artificial representa uma preocupação central em um cenário de profunda insegurança jurídica resultante da célere evolução experimentada sobretudo no domínio das novas tecnologias, com potenciais riscos importantes de vulneração a direitos humanos.

Há algum consenso entre os estudiosos da interação entre Direito e Tecnologia de que os profundos impactos jurídicos resultantes do advento e evolução da Inteligência Artificial (IA) têm sido menosprezados por boa parte da comunidade de "operadores do Direito"[1].

Entretanto, pode-se reconhecer aqui, substancialmente, o mesmo risco identificado (e confirmado) no desenvolvimento da Internet[2] quanto às consequências deletérias da ausência de avaliação de aspectos jurídicos e éticos.

A diferença talvez resida no caráter quase "apocalítico" que muitos identificam nas potencialidades não apenas disruptivas mas verdadeiramente destrutivas de uma "superinteligência artificial" cujas metas possam se desalinhar dos valores humanos centrais[3].

---

[1] A esse propósito, revela-se pertinente a invocação da conhecida Lei de Amara, segundo a qual tende-se a superestimar os efeitos das novas tecnologias no curto prazo e subestimá-las no longo prazo. No original: "*We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run*". Entre outras fontes, pode-se citar: https://www.oxfordreference.com/view/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00018679 [Consult. 03 jul. 2019]

[2] É o que afirma LAWRENCE LESSIG em "*Code and other Laws of Cyberspace*". Na síntese que realizei do pensamento do constitucionalista norte-americano em artigo publicado no Brasil: "De fato, a web, desde o seu início, sempre foi vista como uma espécie de "terra sem-lei", em que as liberdades individuais deveriam prevalecer em detrimento de qualquer tentativa de controle externo. As inúmeras anomalias que brotaram nesse contexto anárquico, entre as quais se poderia destacar toda sorte de crimes praticados sob o manto do anonimato, compeliram o Estado a engendrar mecanismos regulatórios do espaço cibernético para impor alguma ordem ao caos que ameaçava instalar-se. Paralelamente, o próprio Estado viu-se na contingência de se utilizar dessas novas ferramentas para realizar, com maior eficiência e eficácia, os seus fins, passando, deste modo, a padecer das diversas conseqüências da ausência de regulação no universo virtual. Ocorre que até relativamente pouco tempo, o processo de tomada de decisão, quanto aos padrões operacionais que forjaram a atual estrutura da Internet, encontrava-se atribuído, quase que exclusivamente, a técnicos ou especialistas em tecnologia informação. Com efeito, apenas recentemente os operadores do direito despertaram da profunda letargia ou mesmo ojeriza que caracterizava sua relação com novas tecnologias. Consoante ressalta LESSIG, as arquiteturas de regulação que foram erigidas ao longo das últimas décadas ergueram-se de acordo com critérios fundamentalmente de ordem técnica ou operacional determinados sobretudo por dois vetores de força: o Mercado e o Estado". PAULA, Gáudio Ribeiro de - Uma Introdução ao "Direito de Informática", p. 8-9.

[3] Entre outros, BOSTROM apresenta pertinentes preocupações quanto ao desenvolvimento de máquinas super-inteligentes, ressaltando a necessidade de dotá-las de motivações compatíveis com os seres humanos: "*The ethical issues related to the possible future creation of machines with general intellectual capabilities far outstripping those of humans are quite distinct from any ethical problems arising in current automation and information systems. Such superintelligence would not be just another technological development; it would be the most important invention ever made, and would lead to explosive progress in all scientific and technological fields, as the superintelligence would conduct research with superhuman efficiency. To the extent that ethics is a cognitive pursuit, a superintelligence could also easily surpass humans in the quality of its moral thinking. However, it would be up to the designers of the superintelligence to specify its original motivations. Since the superintelligence*

Além da vulnerabilidade já identificada em ferramentas atualmente empregadas, ofensivas a direitos como isonomia[4], devido processo legal, privacidade[5], entre outros, divisam-se virtuais malferimentos de garantias ainda mais fundamentais, como é o caso do próprio direito à vida[6].

De outra parte, os novos recursos tecnológicos estão a amplificar o conhecido problema do ritmo ("*pacing*"), relativo à dificuldade de a legislação acompanhar a rápida evolução dos fatos sociais[7]. O já evidente descompasso entre legislação e problemas jurídicos emergentes das novas tecnologias alcança no caso da IA um patamar inédito, ante o possível surgimento célere de realidades até aqui apenas imaginadas[8].

Sobreleva realçar que a ubiquidade do fenômeno (IA) alcança inclusive os órgãos atualmente responsáveis pelo protagonismo no processo regulatório[9]. Com efeito, mesmo na seara dos entes legiferantes o uso da tecnologia vem sendo cogitado (e, em alguns casos, empregado) com algum entusiasmo para identificar formas mais eficazes de regular

---

may become unstoppably powerful because of its intellectual superiority and the technologies it could develop, it is crucial that it be provided with human-friendly motivations". BOSTROM, Nick - *Ethical Issues in Advanced Artificial Intelligence*, p. 1.

[4] Um dos exemplos que costumam ser citados aqui é o caso *State v. Loomis*, no qual se debateu a suposta conduta anti-isonômica do sistema *COMPAS* ("*Correctional Offender Management Profiling for Alternative Sanctions*"), utilizado por diversos órgãos jurisdicionais norte-americanos como ferramenta de suporte à decisão com o objetivo de avaliar, entre outros aspectos, risco de reincidência criminal. Alegou-se que os algoritmos empregados pela Northpointe (empresa desenvolvedora) faziam uso de dados estatísticos que comprometeriam a necessária isenção e individualização da pena. A síntese do caso pode ser encontrada, dentre outras, em matéria publicada na Harvard Law Review sob o título "*State v. Loomis - Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing*".

[5] Ilustrativo a esse respeito o Sistema de Crédito Social em implementação na China, por meio do qual se estabeleceu uma espécie de "score" para pontuar condutas desejáveis e indesejáveis dos cidadãos chineses a partir da captura de dados por meio de sofisticados sistemas artificialmente inteligentes, que fazem uso de recursos como reconhecimento facial e identificação de outros dados biométricos. Entre as curiosas sanções concebidas para desestimular o mau comportamento encontra-se a restrição de acesso a alguns meios de transporte (v.g. aviões). XU, V. X., & XIAO, B. - *China's social credit system seeks to assign citizens scores, engineer social behaviour* , p. 3.

[6] É o caso dos Sistemas Autônomos de Armas Letais (*Lethal Autonomous Weapons Systems - LAWS*), quanto aos quais já existe algum clamor para sua vedação. Entre outros, pode ser citado o esforço do "*Future of Life Institute*" ("*FLI*"). Informações disponíveis em: https://futureoflife.org/laws-pledge/ [Consult. 28 jun. 2019]. Interessante estudo sobre as implicações bélicas da conjunção entre técnicas de biomimetismo e inteligência artificial pode ser encontrado em: DAMIANO, Jean-Pierre - *Biomimétisme, intelligence artificielle et robotique. Applications de l'intelligence en essaim et questions d'éthique et de droit*.

[7] HAGEMANN, Ryan & SKEES, Jennifer Huddleston & THIERER, Adam – *Soft Law For Hard Problems: The Governance Of Emerging Technologies In An Uncertain Future Intelligence*, p. 37.

[8] Veja-se, exemplificativamente, o caso da imputação de responsabilidade ou mesmo personalidade jurídica a agentes artificialmente inteligentes, que vem sendo objeto de intenso debate doutrinário. A esse respeito: SOLUM, Lawrence B. - *Legal Personhood for Artificial Intelligences*; ANDRADE, Francisco & NOVAIS, Paulo & NEVES, José - *Issues on Intelligent Electronic Agents and Legal Relations; e* WETTIG, Steffen and ZEHENDNER, Eberhard – *A legal analysis of human and electronic agents*.

[9] Em verdade, alcança todos os poderes constituídos, inclusive o Judiciário. A esse propósito importante estudo sobre o tema da substituição de julgadores humanos por máquinas pode ser encontrado em: PEREIRA, Alexandre Libório Dias – *Ius ex Machina*? Da Informática Jurídica ao Computador-Juiz.

determinadas questões, assim como fazer uma espécie de acompanhamento da eficácia e efetividade das normas editadas ou até análises preditivas[10].

Nesse cenário, um dos desafios nos quais se debruçam especialistas de diversos campos do conhecimento consiste no estabelecimento de "barreiras de contenção" ou "redes de segurança" para assegurar a integridade sobretudo de direitos humanos ante o progresso exponencial dos sistemas artificialmente inteligentes[11].

Foi precisamente a partir de tais preocupações que se formularam conceitos como "*human rights by design*"[12], "*beneficial AI*"[13], "*AI for good*"[14] e "*Human-Centered AI*"[15]. Todos evidenciam um esforço para assegurar algum alinhamento entre o desenvolvimento tecnológico e os valores humanos centrais.

Nesse contexto, reputa-se bem-vinda toda investigação que possa oferecer contribuições para a identificação de estratégias regulatórias eficientes para disciplinar referidos sistemas.

É o que se pretende realizar com o presente estudo, a partir de uma análise histórica, com uma perspectiva comparatista e algum aporte multidisciplinar resultante de incursões objetivas em domínios meta-jurídicos.

---

[10] A esse respeito, entre outros, pode ser citado o estudo conduzido por VOERMANS, do Centro de Estudos Legislativo da Universidade de Tilburg, na Holanda. VOERMANS, Wim – *Computer-assisted legislative drafting in the Netherlands: the LEDA-system*.

[11] Para aquilitar os riscos concretos de "acidentes" envolvendo sistemas de IA a partir de uma análise objetiva, mas tecnicamente consistente, veja-se AMODEI, Dario & OLAH, Chris & SCHULMAN, John & STEINHARDT, Jacob & CHRISTIANO, Paul & MANÉ, Dan – *Concrete Problems in AI Safety*.

[12] Trata-se de um modelo proposto por diferentes atores para introduzir travas sistêmicas que restrinjam o desenvolvimento e implementação de agentes artificialmente inteligentes aptos a provocarem lesões a direitos humanos.

[13] Esse último foi o mote de conferência realizada em Asilomar pelo "*Future of Life Institute*" ("*FLI*"). Fonte: https://futureoflife.org/bai-2017/?cn-reloaded=1 [Consult. 25 jun. 2019].

[14] Trata-se de uma plataforma das Nações Unidas que promove o diálogo sobre o uso benéfico da Inteligência Artificial, desenvolvendo projetos concretos. Informações disponíveis em: https://www.itu.int/en/ITU-T/AI/Pages/default.aspx [Consult. 28 jun. 2019].

[15] Essa é a linha terminológica seguida pelo Centro de Estudos de Stanford voltado para a pesquisa e ensino multidisciplinares sobre inteligência artificial. Fonte: https://hai.stanford.edu/ [Consult. 17 jul. 2019].

CAPÍTULO I

1. Inteligência artificial

"*Ubi societas, ibi jus*" afirmava a sabedoria jurídica dos romanos, na clássica formulação de Ulpiano no "*Corpus Juris Civilis*". De fato, onde está a sociedade deve estar o Direito, como condição mesma para a sua subsistência ante o caráter eminentemente conservador do fenômeno jurídico.

Nessa perspectiva, convém indagar para onde ruma a sociedade pós-moderna. A sua fluidez, inspiração para que BAUMAN cunhasse a conhecida expressão "modernidade líquida"[16], trouxe imensos desafios à comunidade jurídica para conceber soluções jurídicas condizentes a uma "sociedade em rede" na "era da informação", para empregar os termos consagrados por CASTELLS[17].

O advento da rede mundial de computadores (Internet)[18] parece ter dado início a mais uma entre tantas revoluções tecnológicas que marcaram os dois últimos séculos, com desdobramentos ubíquos em todos os âmbitos das atividades humanas.

[16] "Fluidez é a qualidade de líquidos e gases. (…) Os líquidos, diferentemente dos sólidos, não mantêm sua forma com facilidade. (…) Os fluidos se movem facilmente. Eles "fluem", "escorrem", "esvaem-se", "respingam", "transbordam", "vazam", "inundam" (…) Essas são razões para considerar "fluidez" ou "liquidez" como metáforas adequadas quando queremos captar a natureza da presente fase (…) na história da modernidade". BAUMAN, Zygmunt - Modernidade líquida, p. 7-9.

[17] CASTELLS, Manuel - A Sociedade em Rede. A Era da Informação: Economia, Sociedade e Cultura - Volume 1.

[18] Na síntese que fiz em modesto artigo sobre o denominado "Direito de Informática", segue breve descrição sobre a criação da Internet: "O surgimento e desenvolvimento da rede mundial de computadores (Internet) encontra-se associado à criação e implementação das tecnologias antes referidas [computador moderno, modem, redes, etc.], as quais puderam oferecer o aporte instrumental necessário à sua viabilização. Assim é que se pode costuma aludir a um primeiro momento, quando a Internet era a então ARPANET, rede implementada nos Estados Unidos, em 1969, pela ARPA (*Advanced Research Projects Agency*), graças a tecnologias já disponíveis como o modem[1] na década de 60. Com objetivos subjacentes de natureza militar, o projeto estabelecia vias de comunicação eficientes entre cientistas, com estrutura de células autônomas, interconectando as Universidades da Califórnia (UCLA) e Stanford, as primeiras universidades interligadas pela rede. O crescimento vertiginoso e a popularização da rede – que deixou, rapidamente, o solo americano, despindo-se dos intuitos bélicos e passando a ostentar propósitos acadêmicos e, posteriormente, comerciais, sociais, governamentais, lúdicos, ... –, deu-se a partir de ferramentas e interfaces que foram agregadas paulatinamente para permitir ao usuário conectar-se e buscar informações de maneira fácil e intuitiva. São exemplos, nesse sentido: a criação do e-mail, em 1971, por TOMLINSON; a adoção do protocolo de comunicação TCP/IP, em 1983 – que permitiu o estabelecimento de regras uniformes para o tráfego de dados pela rede –, graças ao trabalho de KHAN e CERF; a introdução do sistema de nomes de domínios (DNS – *Domain Name Service*), para a criação de endereços de páginas na web, por MOCKAPETRIS, em 1984; a instituição da arquitetura informacional em formato de teia e utilização de hipertextos, a www (*World Wide Web*), com as contribuições do pesquisador inglês BERNERS-LEE, em 1989". PAULA, Gáudio Ribeiro de - Uma Introdução ao "Direito de Informática", p. 3.

Até relativamente pouco tempo, contudo, boa parte das decisões relativas à definição das arquiteturas de regulação de tais inovações era tomada por instâncias eminentemente técnicas e operacionais, sem qualquer ponderação quanto aos aspectos jurídicos[19].

Contudo, a necessidade de instituir marcos regulatórios emerge dos riscos inerentes à instabilidade jurídica que costuma surgir nos momentos de disrupção evolutiva[20], a produzir desequilíbrios e disfunções a comprometer, em muitos casos, a integridade sistêmica do próprio ordenamento jurídico. Pense-se, ilustrativamente, nas intrincadas questões jurídicas em torno dos seguintes temas: a) direito à privacidade e a necessidade de vigilância das informações[21] que circulam nas redes sociais para prevenir atos de terrorismo; b) "fake news" e censura à liberdade de imprensa; c) tributação de operações transnacionais realizadas por meio de "smart contracts" sem intervenção humana; d) imputação de responsabilidade civil e penal em acidentes envolvendo carros autônomos; e e) relações contratuais baseadas em criptomoedas.

A expressão "tecnologia disruptiva" parece ter sido cunhada por BOWER e CHRISTENSEN em artigo publicado na Revista de Negócios de Harvard sob o título "*Disruptive Technologies: Catching the Wave*", que teria sido popularizada a partir do final da década de 1990[22]. Embora tenha sido concebido sob um enfoque econômico, o termo

---

[19] É o que afirma LAWRENCE LESSIG em "Code and other Laws of Cyberspace". LESSIG, Lawrence - *Code and other Laws of Cyberspace*. Na síntese que realizei do pensamento do constitucionalista norte-americano: "De fato, a web, desde o seu início, sempre foi vista como uma espécie de "terra sem-lei", em que as liberdades individuais deveriam prevalecer em detrimento de qualquer tentativa de controle externo. As inúmeras anomalias que brotaram nesse contexto anárquico, entre as quais se poderia destacar toda sorte de crimes praticados sob o manto do anonimato, compeliram o Estado a engendrar mecanismos regulatórios do espaço cibernético para impor alguma ordem ao caos que ameaçava instalar-se. Paralelamente, o próprio Estado viu-se na contingência de se utilizar dessas novas ferramentas para realizar, com maior eficiência e eficácia, os seus fins, passando, deste modo, a padecer das diversas conseqüências da ausência de regulação no universo virtual. Ocorre que até relativamente pouco tempo, o processo de tomada de decisão, quanto aos padrões operacionais que forjaram a atual estrutura da Internet, encontrava-se atribuído, quase que exclusivamente, a técnicos ou especialistas em tecnologia informação. Com efeito, apenas recentemente os operadores do direito despertaram da profunda letargia ou mesmo ojeriza que caracterizava sua relação com novas tecnologias. Consoante ressalta LESSIG, as arquiteturas de regulação que foram erigidas ao longo das últimas décadas ergueram-se de acordo com critérios fundamentalmente de ordem técnica ou operacional determinados sobretudo por dois vetores de força: o Mercado e o Estado". PAULA, Gáudio Ribeiro de - *Op. cit.*, p. 8-9.

[20] Eis o que alerta NATHAN CORTEZ em interessante estudo sobre a regulação de tecnologias disruptivas: "*A persistent challenge for regulators is confronting new technologies or business practices that do not square well with existing regulatory frameworks. These innovations can depart in important ways from older incumbents. For example, the innovation might present unanticipated benefits and risks. It might disturb carefully crafted equilibria between regulators, industry, and consumers. The innovation might puncture prevailing regulatory orthodoxies, forcing regulators to reorient their postures or even rethink their underlying statutory authority. The quintessential example is the Internet, which rumpled not just one, but several regulatory frameworks, including those of the Federal Communications Commission ("FCC"), the Federal Trade Commission ("FTC"), and the Food and Drug Administration ("FDA")*". CORTEZ, Nathan - *Regulating Disruptive Innovation*, p. 176.

[21] Não é demais recordar que a Constituição Portuguesa consagra particular proteção aos dados informáticos em seu art.º 35.º.

[22] *BOWER, Joseph L. & CHRISTENSEN, Clayton M. (1995). "Disruptive Technologies: Catching the Wave" Harvard Business Review, January–February 1995. Apud* CORTEZ, Nathan - *Op. cit.*, p. 177. O texto original

passou a ser empregado de forma mais abrangente para descrever os saltos tecnológicos que provocam a ruptura de modelos ou padrões tradicionais relativos a produtos ou serviços[23].

Em interessante estudo sobre o tema, ROY AMARA indica a existência de quatro fases a marcar a trajetória de adesão a tais tecnologias ("hype cycle")[24]: 1.ª) Gatilho da Tecnologia ("*Technology Trigger*") - momento em que o avanço tecnológico potencial dá seus pontapés iniciais, ocorrendo as primeiras provas de conceito e o interesse da mídia a provocar publicidade significativa, mesmo sem a comprovação de viabilidade comercial; 2.ª) Pico das Expectativas Infladas ("*Peak of Inflated Expectations*") - a publicidade antecipada produz algumas histórias de sucesso alimentado as elevadas expectativas em torno da novidade tecnológica, sem considerar os diversos fracassos já experimentados em profusão; 3.ª) Vale da Desilusão ("*Trough of Disillusionment*") - o interesse diminui à medida que os resultados prometidos não são alcançados, de modo que os investimentos passam a se concentrar nos provedores sobreviventes aptos a melhorarem seus produtos para satisfazer os interesses dos primeiros adotantes ("*early adopters*"); 4.ª) Aclive da Iluminação ("*Slope of Enlightenment*") - avanços começam a se consolidar com exemplos de benefícios concretos, tornando-se mais amplamente compreendidos, embora empresas conservadoras permaneçam cautelosas; e 5.ª) Planalto da Produtividade "*Plateau of Productivity*" - a adesão generalizada começa a decolar, uma vez que os critérios para avaliar a viabilidade do provedor são mais claramente definidos[25].

Com base em tal modelo, a empresa de consultoria Gartner apresentou o seguinte ciclo de expectativas de tecnologias emergentes para o ano de 2017:

pode ser encontrado em https://hbr.org/1995/01/disruptive-technologies-catching-the-wave [Consult. 27 mai. 2018].

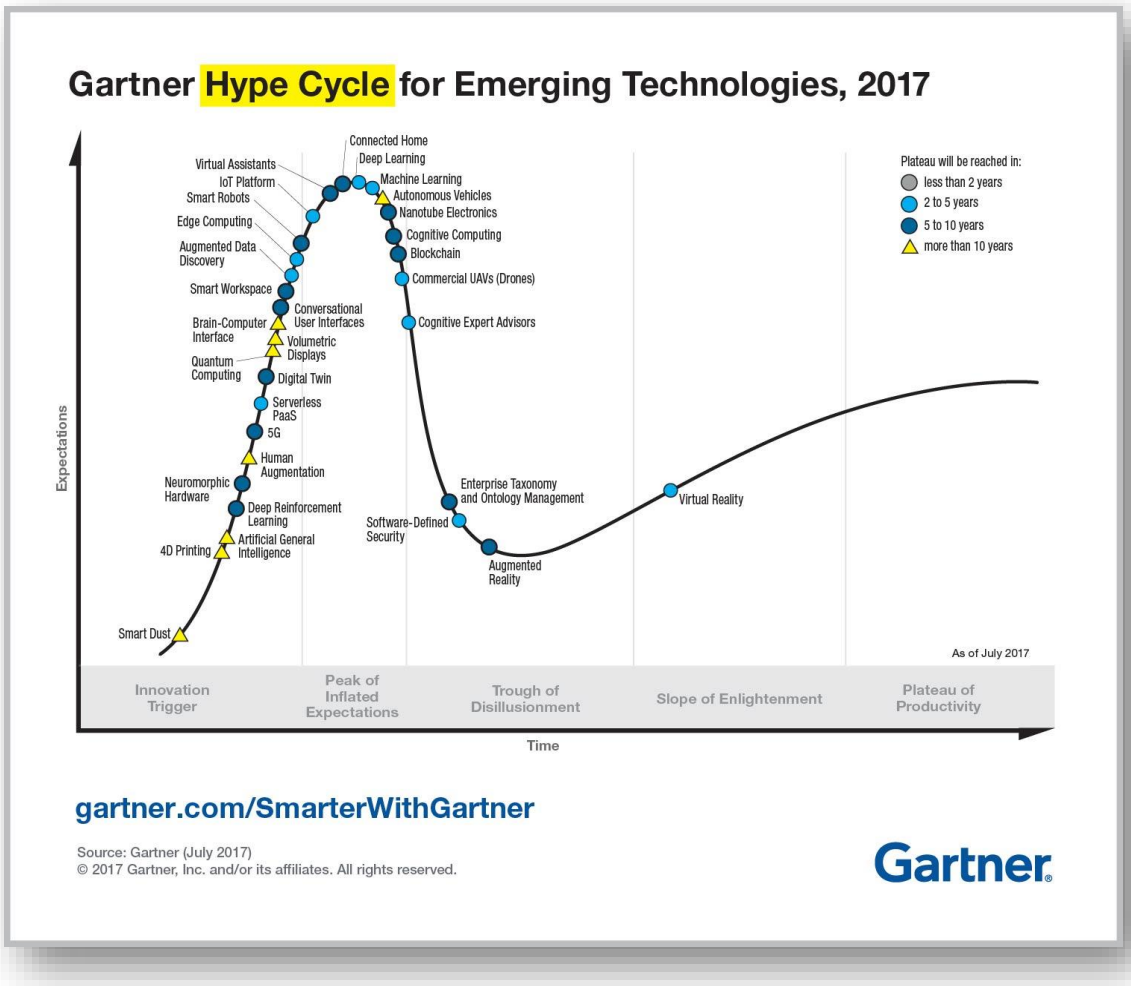
[23] Impressão tridimensional, realidade virtual, criptomoedas, inteligência artificial e automação veicular seriam alguns dos inúmeros exemplos.

[24] A teoria passou a ser denominada "Amara´s Law". Ex-Presidente do Instituto para o Futuro ("*Institute for the Future"* - http://www.iftf.org), Roy Amara foi um pesquisador e cientista (1925-2007) recordava que se tende a superestimar o efeito das tecnologias no curto prazo e subestimá-los no longo prazo: "*We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run". https://www.oxfordreference.com/view/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00018679* [Consult. 27 mai. 2018].

[25] Síntese disponível em https://en.wikipedia.org/wiki/Hype_cycle [Consult. 27 mai. 2018].

Figura 1.

Ciclo de "entusiasmo" quanto a tecnologias emergentes [26]



As consequências da implementação de tais inovações disruptivas não são, naturalmente, previsíveis, mas há sinalizações de que são justificáveis os receios de uma pequena parte da comunidade jurídica ocupada em promover estudos de uma sorte de "futurologia jurídica".

Embora os avanços tecnológicos possam resultar em notáveis progressos na qualidade de vida e no desenvolvimento econômico[27], evidentemente, o Estado não deveria ficar inteiramente alheio às questões jurídicas que possam emergir de tais novidades.

[26] Disponível em https://blogs.gartner.com/smarterwithgartner/files/2017/08/Emerging-Technology-Hype-Cycle-for-2017_Infographic_R6A.jpg [Consult. 27 mai. 2018].

[27] Em rico estudo sobre o tema, o Departamento de Economia e Assuntos Sociais da Organização das Nações Unidas, assim conclui: "*Technological progress is a main driver of economic growth and improvements in living standards. It increases productivity, thereby boosting per capita income and consumption. Technology also influences the nature and quality of work, as well as the structure of societies. Of note, past literature has shown that technology, institutions and society tend to evolve together*". ORGANIZAÇÃO DAS NAÇÕES UNIDAS - *The impact of the technological revolution on labor markets and income distribution*, p. 4.

Contudo, a intervenção heterônoma estatal tem se mostrado, em diversos casos, desastrosa. O fracasso decorre, em regra, da tentativa de empregar soluções regulatórias tradicionais para situações que escapam aos atuais institutos e conceitos jurídicos incapazes de garantir o necessário equilíbrio entre o interesse público e o estímulo à inovação.

No tocante à inteligência artificial (IA), cuja evolução exponencial foi marcante na última década, entre as incontáveis possibilidades trazidas pelo seu emprego no terreno jurídico, no campo da formação dos contratos, a implementação dos chamados "smart contracts" (contratos inteligentes autoexecutáveis) ou "*e-contracts*"[28] permite a automação inteligente de operações resultantes de negócios jurídicos[29].

Tais contratos eletrônicos firmados sem qualquer intervenção humana recebem em Portugal (assim como em diversos outros países) o mesmo tratamento jurídico, essencialmente, destinado aos contratos convencionais[30].

Contudo, "como aplicar o Direito Contratual a negócios firmados por agentes autônomos sem atribuir-lhes intencionalidade?", indagam alguns doutrinadores. E prosseguem: "Devemos entender que agentes autônomos têm intenções? Ou devemos reconstruir o Direito Civil sem o conceito de intencionalidade?"[31]. Essas seriam algumas das questões que evidenciam a precariedade da resposta normativa estatal.

---

[28] "Contratos inteligentes são escritos como código de programação que pode ser executados em um computador, em vez de um documento impresso com uma linguagem legal. Este código pode definir regras estritas e consequências da mesma forma que um documento legal tradicional, estabelecendo as obrigações, benefícios e penalidades que podem ser devidas a qualquer das partes em várias circunstâncias diferentes. Mas, ao contrário de um contrato tradicional, ele também pode tomar informações como uma produção de outros bens, processar essa informação através das regras estabelecidas no contrato e tomar quaisquer medidas necessárias dele como resultado". Extraído de https://guiadobitcoin.com.br/um-guia-para-iniciantes-sobre-smart-contracts/ [Consult. 13 mai. 2018].

[29] A prevenção de violações a direitos autorais ou a propriedade intelectual representaria uma das tantas aplicações dos "smart contracts" no que se convencionou denominar "Digital Rights Management" (gestão ou gerenciamento de direitos digitais) ou "DRM". A partir dessa modalidade contratual não seria processados determinados comandos para impedir a cópia ou transferência de arquivos digitais de música ou vídeo protegidos por direitos autorais. Extraído de https://guiadobitcoin.com.br/um-guia-para-iniciantes-sobre-smart-contracts/ [Consult. 13 mai. 2018].

[30] PORTUGAL, Decreto-Lei 7/2004 (em atenção à Diretiva 2000/31) - Art.º 33.º Contratação sem intervenção humana 1. À contratação celebrada exclusivamente por meio de computadores, sem intervenção humana, é aplicável o regime comum, salvo quando este pressupuser uma actuação. 2. São aplicáveis as disposições sobre erro: a) Na formação da vontade, se houver erro de programação; b) Na declaração, se houver defeito de funcionamento da máquina; c) Na transmissão, se a mensagem chegar deformada ao seu destino. 3. A outra parte não pode opor-se à impugnação por erro sempre que lhe fosse exigível que dele se apercebesse, nomeadamente pelo uso de dispositivos de detecção de erros de introdução.

[31] Entre outros, VIEIRA, Miguel Marques - A autonomia privada na contratação electrônica sem intervenção humana; e MARANHÃO, Juliano - A pesquisa em inteligência artificial e Direito no Brasil. NATALINO IRTI, citado pelo primeiro, chega a afirmar que tais contratos eletronicos prescindiriam de qualquer manifestação de vontade, pois seriam travados "sem diálogo" ou "em silêncio". *Apud* VIEIRA, Miguel Marques - *Op. cit.*, p. 188.

No âmbito específico da proteção aos direitos humanos, vislumbram-se desafios ainda mais ingentes[32].

A partir do cenário europeu, à luz do estudo desenvolvido pelo Comitê *Ad Hoc* Sobre a Inteligência Artificial vinculado ao Conselho da Europa, ao menos quatro "famílias" de direitos humanos[33] sofreriam possíveis impactos ante o desenvolvimento e implementação inadequados de sistemas de IA: i) respeito à dignidade humana; ii) liberdade individual; iii) igualdade, não discriminação e solidariedade; e iv) direitos sociais e econômicos[34].

Além disso, o próprio regime democrático e o Estado de Direito poderiam ser eventualmente comprometidos[35].

Há dois pontos de particular relevo que representam dificuldades inerentes a um significativo conjunto de agentes de IA.

De um lado, a opacidade ("*the black box problem*" [36]) intrínseca à maior parte dos sistemas em desenvolvimento acarreta uma dificuldade particular na identificação de soluções para promover a auditagem de tais tecnologias[37]. A transparência tenderia a evitar, ilustrativamente, o enviazamento algorítmico ("*algorithmic bias*"[38]), a comprometer direitos como isonomia, privacidade, acesso ao direito e tutela jurisdicionais efetiva, devido processo legal, liberdade, autonomia (ou "autodeterminação"), entre outros.

De outro, o deslocamento do eixo epistêmicos das cadeias causais para os supostos "nexos" meramente correlacionais, comumente perceptíveis nos modelos de IA atualmente mais empregados, pode conduzir a graves distorções nas informações geradas e nas decisões tomadas a partir delas[39] (v.g. concessão de crédito, análise de risco para contratação de seguro

---

[32] Eis o que afirmou DUNJA MIJATOVIĆ, Comissário para Direitos Humanos do Conselho Europeu, a propósito: "*Ensuring that human rights are strengthened and not undermined by artificial intelligence is one of the key factors that will define the world we live in*" (Conselho Europeu – *Unboxing artificial intelligence: 10 steps to protect human rights*). Vale notar que, diante da tais preocupações, foi instituído o Comitê Ad Hoc sobre Inteligência Artificial (CAHAI - Ad hoc Committee on Artificial Intelligence), vinculado ao Comitê de Ministros da União Europeia.

[33] Ao abrigo da Convenção Europeia de Direitos Humanos.

[34] CAHAI – *Ad hoc Committee on Artificial Intelligence - Towards Regulation Of AI Systems*, pp. 23-24.

[35] CAHAI – *Op. Cit.*, p. 24.

[36] ZEDNIK, Carlos - *Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence*.

[37] Cuida-se de tema que não passou desapercebido aos constitucionalistas. CANOTILHO, ilustrativamente, assim trata do tema "Ao enfrentar-se os problemas colocados pela inteligência artificial (IA) parece incontornável o grande desafio dos humanos perante a míngua de transparência referente a todos os sistemas desta inteligência. Fala-se em 'caixa preta' que perturba decisivamente o humano e o estado de direito, invocando-se a necessidade de um 'direito subjetivo' à explicação" (*subjective right to explanation*) (WISCHMEYER, 20180". CANOTILHO, José Joaquim Gomes – Sobre a indispensabilidade de uma Carta de Direitos Fundamentais Digitais da União Europeia, p. 69.

[38] DANKS, David & LONDON, Alex John - *Algorithmic Bias in Autonomous Systems*.

[39] O próprio direito de ser submetido a "julgamento" realizado por ser humano tem sido articulado por alguns como direito fundamental. Nesse sentido pode ser citado o art.º 22.º da GDPR europeia (diversamente da LGPD

ou plano de saúde, processo seletivo para emprego ou avaliação de performance ou produtividade para efeito de despedimento por iniciativa do empregador, ou mesmo a prolação de sentença judicial lastreada em dados fornecidos por sistema de IA[40]). Daí resultariam potencialmente malferidas garantias como integridade pessoal, saúde, liberdade de escolha de profissão e segurança no emprego.

Relativamente às repercussões quanto aos direitos humanos, podem ser mencionados três aspectos a serem considerados na regulação da matéria: i) os direitos de um indivíduo só poderiam ser transferidos com o seu consentimento (princípio de consentimento); ii) a única justificativa para o uso do poder contra a vontade de uma pessoa seria a prevenção de danos (princípio do dano); e iii) o uso da força deveria ser proporcional à ameaça (princípio da proporcionalidade)[41].

O desenvolvimento de uma arquitetura normativa apropriada[42] deve resultar de um esforço mais detido para compreender, de modo abrangente, as implicações jurídicas da adoção de algumas soluções tecnológicas, à luz de um conhecimento interdisciplinar.

Nesse contexto, podem ser mencionadas as seguintes posturas (não necessariamente excludentes entre si) dos agentes reguladores[43]: a) "*wait-and-see*" ("espere e veja") – comportamento passivo em que se aguardam os desdobramentos da tecnologia, enquanto se segue aplicando a legislação em vigor; b) "*adaptive regulation*" ("regulação adaptativa") – representa o abandono ou superação do paradigma "*regulate and forget*" ("regule e esqueça") em direção a um modelo cujas características centrais seriam a responsividade, interatividade e agilidade na produção de normas que se adaptem rapidamente às mudanças tecnológicas[44]; c) "*regulatory sandboxes*" ("caixas de areia regulatórias") – "conjunto de regras que permitiria aos inovadores testar seus produtos e modelos de negócio em um ambiente que

brasileira, cujo dispositivo homólogo - § 3º do art.º 20.º - findou por ser vetado): CASONATO, Carlo - *Intelligenza artificiale e diritto costituzionale: prime considerazioni*.

[40] Uma análise desse ponto quanto ao uso de sistemas de IA na justiça criminal encontra-se em: ALTMAN, Micah & WOOD, Alexandra & VAYENA, Effy - *A HarmReduction Framework for Algorithmic Fairness*.

[41] KRIEBITZ, Alexander & LÜTGE, Christoph - *Artificial Intelligence and Human Rights: A Business Ethical Assessment*, p. 87.

[42] A propósito desse esforço de aprimoramento regulatório geral, digno de nota a iniciativa da União Europeia de criar o programa para a adequação e a eficácia da regulamentação (REFIT). Informações disponíveis em: https://ec.europa.eu/info/law/law-making-process/evaluating-and-improving-existing-laws/refit-making-eu-law-simpler-less-costly-and-future-proof_pt [Acesso em 11/07/20]

[43] EGGERS, William D. & TURLEY, Mike & KISHNANI, Pankaj – *The future of regulationPrinciples for regulating emerging technologies*, p. 11.

[44] Aqui são empregadas técnicas como "*trial and error*" ("tentativa e erro"), "*co-design of regulation*" ("regulação participativa") "feedback loops" ("retroalimentação circular"). Instrumentos comumente utilizados seriam os mecanismos de "*soft law*", para criar expectativas jurídicas substanciais, mas não diretamente coercíveis. EGGERS, William D. & TURLEY, Mike & KISHNANI, Pankaj – *Op. Cit.*, p. 11.

temporariamente os isenta de algum e de vários requisitos legais"[45]; e d) "*collaborative regulation*" ("regulação colaborativa") – em que o alinhamento regulatório nacional e internacional ocorre por meio do engajamento dos maiores envolvidos ("*players*") no ecossistema tecnológico[46].

Seja como for, a construção e implementação de um infra-estrutura regulatória revela-se essencial, ademais, para criar uma cadeia de confiança entre desenvolvedores e consumidores de tais tecnologias[47].

O equilíbrio que se intenciona alcançar poderia ser assim resumido, nas palavras de SEOANE: "*Esperemos que nuestro pensamiento jurídico, actual y futuro, junto a los ingenieros, físicos, matemáticos y demás científicos, sea capaz de encontrar fórmulas para disciplinar las fuerzas tecnológicas y sus rendimentos, sin entorpecer su avance pero garantizando al miesmo tiempo la pervivencia de los derechos de las personas (humanas: valga aquí el pleonasmo)*"[48].

O objetivo central da investigação a ser realizada consiste, precisamente, na coleta e análise de informações que permitam realizar uma avaliação das estratégias regulatórias mais adequadas para assegurar a conciliação entre o estímulo ao desenvolvimento de sistema de IA e a proteção contra potenciais agressões a direitos humanos fundamentais.

Seu centro de gravidade consistirá na análise de modelo proposto por diversas instituições acadêmicas, organismos internacionais e mesmo agências regulatórias estatais, consistente na introdução de "travas sistêmicas" que, por meio de restrições intrínsecas, obstem a criação, desenvolvimento e implementação de agentes artificialmente inteligentes que possam causar violações a direitos humanos.

Trata-se da abordagem conhecida como "*human rights by design*", cujas possibilidades e limites regulatórios figuram com cerne da dissertação.

---

[45] EGGERS, William D. & TURLEY, Mike & KISHNANI, Pankaj – *Op. Cit*., p. 12.
[46] EGGERS, William D. & TURLEY, Mike & KISHNANI, Pankaj – *Op. Cit*., p. 17.
[47] No cenário europeu, encontra-se em curso projeto bastante audacioso e abrangente nessa direção, consoante ressalta a Comissão Europeia em importante documento sobre o tema: "*The key elements of a future regulatory framework for AI in Europe that will create a unique 'ecosystem of trust'. To do so, it must ensure compliance with EU rules, including the rules protecting fundamental rights and consumers' rights, in particular for AI systems operated in the EU that pose a high risk7 . Building an ecosystem of trust is a policy objective in itself, and should give citizens the confidence to take up AI applications and give companies and public organisations the legal certainty to innovate using AI. The Commission strongly supports a human-centric approach based on the Communication on Building Trust in Human-Centric AI8 and will also take into account the input obtained during the piloting phase of the Ethics Guidelines prepared by the High-Level Expert Group on AI*". Comissão Europeia - *White Paper on Artificial Intelligence - A European approach to excellence and trust*.
[48] SEOANE, A. MOZO - *La revolución tecnológica y sus retos: medios de control, falos de los sistemas y ciberdelincuencia*, pág. 98.

Consiste em estratégia de regulação qualificada como tecnoregulatória[49], isto é, uma forma de regular a conduta por meio da própria tecnologia. Cuida-se da normatização do comportamento humano a partir do uso de aparatos ("artefatos") tecnológicos nos quais se encontram inscritas regras de conduta e com os quais se poderia até mesmo inviabilizar (ou dificultar sobremaneira) fisicamente o descumprimento de tais normas.

Conceito próximo seria o de "captologia" ("*captology*"), que poderia ser considerado uma variação da tecno-regulação. Cuida-se de uma espécie de "ciência" cuja ênfase recairia sobre o "*design*", pesquisa, e análise de produtos computacionais interativos desenvolvidos com o objetivo mudar atitudes comportamentais. Estuda, pois, todos os modos de influenciar intencionalmente a conduta humana por meio de tais instrumentos[50].

A depender do contexto, os modelos regulatórios desse jaez podem ser referidos a partir das locuções "regulação pela tecnologia" ("*regulation by technology*"), "normatividade tecnológica" ("t*echnological normativity*"), "programas reguladores" ("*regulative software*"), "Direito como design" ("*law as design*"), "regulação baseada em design" ("*design-based regulation*") ou "regulação algorítimica" ("*algorithmic regulation*")[51].

Não se ignora, de outra parte, o importante desafio concernente à formação de consenso internacional[52] em torno do conjunto de direitos humanos a serem tutelados por meio de tal perspectiva regulatória e do modo como deveriam ser interpretados e reconhecidos, na contraposição entre universalismo e relativismo cultural[53].

Entretanto, não se pode taxar o projeto de utópico e desprovido da possibilidade de concreção. Além dos esforços empreendidos quanto à definição de princípios ("*soft law*") e

[49] BAYAMLIOĞLU e LEENES asseveram que o neologismo "tecno-regulação" (ou "*techno-regulation*") traduziria a "influência intencional do comportamento dos indivíduos ao incorporar normas nos sistemas e dispositivos tecnológicos". BAYAMLIOĞLU, Emre & LEENES, Ronald - *The 'rule of law' implications of data-driven decision- making: a techno-regulatory perspective*, p. 298.

[50] LEENES, Ronald E - *Framing Techno-Regulation: An Exploration of State and Non-State Regulation by Technology*, p. 154.

[51] BAYAMLIOĞLU, Emre & LEENES, Ronald – *Op. cit.*, p. 298.

[52] Cuida-se de premissa essencial à proteção universalmente ampla a ser assegurada aos direitos humanos, considerada a já mencionada ubiquidade dos sistemas de inteligência artificial, particularmente presente na denominada "inteligência artificial distribuída" ("*Distributed Artificial Intelligence – DAI*"). Uma síntese didática sobre o assunto pode ser enco ntrada em: https://francesco-ai.medium.com/distributed-artificial-intelligence-3e3491e0771c [Acesso em 24/12/20].

[53] Entre outros, o tema encontra-se bem explorado em: KROETZ, Flávia Saldanha - *Between global consensus and local deviation: a critical approach on the universality of human rights, regional human rights systems and cultural diversity*.

regras ("*hard law*") por entidades privadas[54], organismos internacionais[55], revela-se ilustrativa a tentativa promovida por diversos países de criar normas diretamente voltadas para as máquinas ("*machine-consumable legislation*")[56] que consistiriam em uma espécie de tecno-regulação[57], mediante o emprego de leis traduzidas em códigos a serem interpretados diretamente pelos sistemas de IA. Ainda em fase experimental, o modelo vem sendo testado na Austrália, Nova Zelândia e Inglaterra[58].

Corroboram essa perspectiva de enfrentamento do problema regulatório as ponderações de CANOTILHO, segundo o qual "o cerne problemático da revolução algorítmica não está na 'regulação' da IA ancorada apenas em direitos fundamentais, mas no desenvolvimento objectivo de estruturas organizacionais e mecanismos procedurais que permitam um efectivo controlo pelas autoridades competentes e tribunais com mecanismos adequados à salvaguarda do sujeito colocado no princípio e no fim do sistema"[59].

Sobreleva notar, por fim, que, embora especialmente intrincada, a necessária interação entre tecnologia e ciências humanas pode produzir uma notável evolução nas abordagens teóricas e práticas da IA, sobretudo no tocante ao que PIERRE LEVY qualifica

---

[54] Uma das tantas iniciativas consiste na abrangente proposta de uso de *soft law* para regular o fenômeno da IA conduzida pelo Instituto de Engenharia Elétrica e Eletrônica (IEEE), de que resultou o lançamento da "Iniciativa Global do IEEE sobre Ética de Sistemas Autonomos e Inteligentes" ("*IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*"). O inteiro teor encontra-se disponível em: https://standards.ieee.org/industry-connections/ec/autonomous-systems.html [Consult. 8 jan. 2020]

[55] Um dos tantos exemplos que poderiam ser recordados é o da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), que produziu importante instrumento de "soft law", a Recomendação sobre Inteligência Artificial, aprovada em 22 de maio de 2019 pelo Conselho Ministerial, a partir de proposta formulada pelo Comitê de Política Econômica Digital (CPED). A íntegra do documento encontra-se disponível em inglês e francês em https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 [Consult. 8 jul. 2019].

[56] WADDINGTON, Matthew – *Machine-consumable legislation: A legislative drafter's perspective – human v artificial intelligence*.

[57] BAYAMLIOĞLU e LEENES descrevem o neologismo "tecno-regulação" (ou "techno-regulation") como a "influência intencional do comportamento dos indivíduos ao incorporar normas nos sistemas e dispositivos tecnológicos" [tradução livre] (BAYAMLIOĞLU, Emre & LEENES, Ronald - The 'rule of law' implications of data-driven decision- making: a techno-regulatory perspective, p. 298). Cuida-se, conforme destaquei em artigo pendente de publicação sobre o tema, da "regulação do comportamento humano a partir do uso de aparatos ("artefatos") tecnológicos nos quais se encontram inscritas regras de conduta e com os quais se poderia até mesmo inviabilizar (ou dificultar sobremaneira) fisicamente o descumprimento de tais normas" (PAULA, Gáudio Ribeiro de Paula – Tecno-Regulação, Inteligência Artificial e Estado de Direito Democrático, p. 8). LEENES ressalta o problema da legitimidade por déficit de transparência em semelhante abordagem regulatória quando se trata de iniciativa estatal: "[...] *state authored techno-regulation has to be seen as supplemental to ³regular´ regulation because legitimacy requires the norms to be transparent and the regulator to be accountable for the norms*" (LEENES, Ronald E – *Framing Techno-Regulation: An Exploration of State and Non-State Regulation by Technology*, p. 143).

[58] WADDINGTON, Matthew – *Op. cit., p. 34*.

[59]CANOTILHO, José Joaquim Gomes – Sobre a indispensabilidade de uma Carta de Direitos Fundamentais Digitais da União Europeia, p. 69.

como "complexidade contextual da significação" ("*complexité contextuelle de la signification*")[60].

1.1. Elementos conceituais

O conceito de inteligência está entre os mais controversos em todas as várias ciências que se debruçam sobre ele. Para o cientista computacional MCCARTHY, seria a parte computacional da habilidade de atingir metas no mundo[61]. Em uma versão mais sintética, MAX TEGMARK a define como a capacidade de resolver problemas complexos (ou "atingir metas complexas")[62].

A maior parte das tentativas de definição da IA tem em comum a associação inevitável à inteligência humana, mesmo que forma indireta e implícita. Parte-se do modelo de funcionamento do cérebro ou mente humanos para se estender a outros "seres" (animais, máquinas, sistemas, etc.) aquelas que seriam características intelectivas supostamente inatas ao homem[63].

Parte dos cientistas da computação a identificam como o campo de estudos sobre "agentes inteligentes"[64], descritos como dispositivos que percebem o ambiente exterior e praticam ações com o objetivo de maximizar as chances de sucessos no cumprimento de seus objetivos[65].

Segundo MACCARTHY, não seria ainda possível identificar que tipos de procedimentos computacionais poderiam ser qualificados como propriamente inteligentes, porquanto apenas parte dos mecanismos da inteligência seria compreendida. Alguns desses

---

[60] Tratar-se-ia de uma via de mão dupla que poderia beneficiar tanto a inteligência artificial quanto as ciências humanas, segundo destaca LEVY: "*En somme, le grand projet de la sémantique computationnelle consiste à construire un pont entre l'ingénierie logicielle et les sciences humaines de telle sorte que ces dernières puissent utiliser à leur service la puissance computationnelle de l'informatique et que celle-ci parvienne à intégrer la finesse herméneutique et la complexité contextuelle des sciences humaines. Mais une intelligence artificielle grande ouverte aux sciences humaines et capable de calculer la complexité du sens ne serait justement plus l'intelligence artificielle que nous connaissons aujourd'hui. Quant à des sciences humaines qui se doteraient d'un métalangage calculable, qui mobiliseraient l'intelligence collective et qui maîtriseraient enfin le médium algorithmique, elles ne ressembleraient plus aux sciences humaines que nous connaissons depuis le XVIIIe siècle : nous aurions franchi le seuil d'une nouvelle épistémè*". LEVY, Pierre - *Intelligence artificielle et sciences humaines*, p. 9.

[61] MCCARTHY, John - *What is Artificial Intelligence*?, p. 2.

[62] TEGMARK, Max - *Life 3.0: Being Human in the Age of Artificial Intelligence*, p. 71.

[63] MCCARTHY, John – *Op. cit.*, p. 3.

[64] TEGMARK oferece definição mais aberta ao afirmar que os agentes inteligentes seriam "[...] *entities that collect information about the environment from sensors and then process the information to decide how to act back on their environment*". TEGMARK, Max - *Life 3.0: Being Human in the Age of Artificial Intelligence*, p. 37.

[65] POOLE, David; MACKWORTH, Alan & GOEBEL, Randy (1998) - *Op. cit.*, p. 1.

mecanismos puderam ser implementados em máquinas e, quanto a esses, sua performance superaria a da inteligência humana[66].

KAPLAN e HAENLEIN definem inteligência artificial como "a capacidade do sistema de interpretar corretamente os dados externos, aprender com esses dados e usar esses aprendizados para atingir metas e tarefas específicas por meio de adaptação flexível"[67]. Do conceito podem ser extraídas as seguintes características: a) habilidade de interpretação de elementos do mundo exterior; b) capacidade de aprendizado; c) adaptabilidade às variações de contexto; d) e comprometimento com o atingimento de metas.

O processo de transferência de competências intelectivas a sistemas computacionais[68] envolve, em algum nível, a simulação de capacidades cognitivas dos homens, mas não se esgota nessa única abordagem metodológica[69].

Sob o manto da locução "Inteligência Artificial" (IA) encontram-se ciências e técnicas (matemática, estatística e ciência da computação) capazes de processar dados para projetar tarefas de processamento de computador complexas[70].

De acordo com um dos seus fundadores[71], o já citado JOHN MCCARTHY, seu objetivo seria o de criar máquinas inteligentes, especialmente a partir de programas de computador desenvolvidos de acordo com uma ampla compreensão da inteligência humana, não confinada por métodos biologicamente observáveis[72].

Sob a perspectiva funcional, KAPLAN e HAENLEIN a definem como a habilidade de um sistema de interpretar corretamente dados externos, aprender a partir deles e utilizar tal aprendizado para atingir metas específicas e tarefas, por meio de adaptação flexível[73].

---

[66] MCCARTHY, John – *Op. cit.*, p. 4.

[67] No original: "*a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation*" KAPLAN, Andreas & HAENLEIN, Michael - *Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence*, p. 15.

[68] Aqui cumpre mencionar a distinção que muitos propõem, a partir dos aportes teóricos críticos de SEARLE, entre IA Fraca ("*Weak AI*") e IA Forte ("*Strong AI*"). A primeira (IA Fraca) se resumiria a uma ferramenta de resolução de problemas (consciência funcional, na terminalogia de SEARLE), enquanto a segunda (IA Forte) consistiria na geração de uma mente artificial (consciência fenomênica), semelhante à humana, relativamente aos seus estados cognitivos. FLOWERS, Johnathan Charles - *Strong and Weak AI: Deweyan Considerations*, p. 1.

[69] *Id*., *ibid*.

[70] RONSIN, Xavier - *In-depth study on the use of AI in judicial systems, notably AI applications processing judicial decisions and data*, p. 31.

[71] Não se pode deixar de registrar o papel essencial das contribuições de ALAN MATHISON TURING, célebre matemático inglês e um dos pais da ciência computacional. Em texto seminal publicado em 1950, TURING formulou a possibilidade teórica de engendrar máquinas que pudessem pensar e enunciou os termos do que viria a se tornar o primeiro teste para avaliar tal capacidade intelectiva na forma de um "jogo de imitação" ("imitation game"). TURING, Alan Mathison - *Computing Machinery and Intelligence*.

[72] MCCARTHY, John - *What is Artificial Intelligence*?, p. 2.

[73] No original: "*a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation*" KAPLAN, Andreas & HAENLEIN,

Um dos conceitos centrais para compreender os sistemas artificialmente inteligentes é o de algoritmo, que poderia ser definido como uma sequência finita de regras formais (operações lógicas e instruções) destinada a tornar possível obter um resultado a partir de um conjunto de dados informados. A sequência poderia ser parte de um processo de execução automática e embutida em modelos desenhados por meio de aprendizado de máquina[74].

Os algoritmos seriam como que o "centro de gravidade cognitivo" dos sistemas de IA. Quanto maior sua capacidade de processamento de dados e geração de informações precisas e relevantes, maior sua eficiência epistêmica.

Um dos recursos fundamentais para a otimização dos algoritmos é o aprendizado de máquina ("*machine learning*"), por meio do qual o sistema se aprimora a partir de sua experiência passada, sem ter sido explicitamente programado para assimilá-la de modo específico[75].

A sua mais conhecida abordagem metodológica denomina-se aprendizagem profunda ("*deep learning*"), a qual faz uso de grande volume de dados tratados por diversas camadas de redes neurais artificiais ("algoritmos inspirados na estrutura de neurônios do cérebro") para resolver problemas de maior complexidade cognitiva (v.g. reconhecimento de objetos em imagens)[76].

1.2. Trajetória histórica

Há alguma controvérsia sobre a data precisa em que o termo "inteligência artificial" (IA ou "*artificial intelligence*" ou AI, para usar a expressão consagrada em inglês[77]) teria

Michael - *Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence*, p. 15.

[74] GLOSSARY - *Apud* RONSIN, Xavier - *In-depth study on the use of AI in judicial systems, notably AI applications processing judicial decisions and data*, p. 31.

[75] https://www.expertsystem.com/machine-learning-definition/

[76] https://machinelearningmastery.com/what-is-deep-learning/

[77] Há quem prefira a expressão inteligência computacional ("*computational intelligence*"), como é o caso de POOLE, para quem a expressão "artificial" sugeriria uma oposição a "real". Em suas reflexões: "*Artificial intelligence (AI) is the established name for the field we have defined as computational intelligence (CI), but the term "artificial intelligence" is a source of much confusion. Is artificial intelligence real intelligence? Perhaps not, just as an artificial pearl is a fake pearl, not a real pearl. "Synthetic intelligence" might be a better name, since, after all, a synthetic pearl may not be a natural pearl but it is a real pearl. However, since we claimed that the central scientific goal is to understand both natural and artificial (or synthetic) systems, we prefer the name "computational intelligence." It also has the advantage of making the computational hypothesis explicit in the name*". POOLE, David; MACKWORTH, Alan & GOEBEL, Randy (1998) - *Computational Intelligence: A Logical Approach*, pp. 1-2.

sido forjada. Enquanto alguns defendem que o termo teria sido cunhado em 1958 por JOHN MCCARTHY, da Universidade de Stanford, outros sustentam que teria surgido em 1956[78].

Um dos responsáveis pela sua gênese científica foi ALAN MATHISON TURING[79], célebre matemático inglês e um dos pais da ciência computacional. Em texto seminal publicado em 1950[80], TURING formulou a possibilidade teórica de engendrar máquinas que pudessem pensar e enunciou os termos do que viria a se tornar o primeiro teste para avaliar tal capacidade intelectiva na forma de um "jogo de imitação" ("imitation game")[81].

A primeira fase (também conhecida como "*the golden years*") da trajetória histórica da AI, que compreende o surgimento do seu próprio conceito, iniciou-se em 1956 com a Conferência de Dartmouth (*Dartmouth Summer Research Project on Artificial Intelligence* – DSRPAI)[82] e encerrou-se em 1974, quando foram suprimidas as verbas destinadas sobretudo pelo governo norte-americano às pesquisas na área em virtude da ausência de resultados concretos mais efetivos ou, para ser mais preciso, da frustração das expectativas criadas[83]. Entre as causas do aparente fracasso das primeiras tentativas de desenvolvimento da IA[84] podem ser lembradas a deficiência técnica dos computadores disponíveis e o alto custo de manutenção de tais equipamentos[85].

---

[78] SUSSKIND, Richared - *Expert Systems in Law: a Jurisprudential Approach to Artificial Intelligence and Legal Reasoning*, p. 171.

[79] Antes de TURING, há quem indique VANNEVAR BUSH como o seu precursor, graças ao seu texto intitulado "As We May Think", em que propôs um sistema que amplificaria o conhecimento e a compreensão humanos. SMITH, Chris; MCGUIRE, Brian; HUANG, Ting & YANG, Gary – *The History of Artificial Intelligence*, p. 4.

[80] Eis como o cientista inaugura suas ponderações: "*I propose to consider the question, ' Can machines think ? ' This should begin with definitions of the meaning of the terms ' machine ' and' think '. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words ' machine' and' think ' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, ' Can machines think ?' is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words*". TURING, Alan Mathison - *Computing Machinery and Intelligence*, p. 433.

[81] Trata-se, em essência, do conhecido teste ("*Turing test*") segundo o qual se poderia reputar inteligente a máquina que pudesse se passar por um humano em uma conversação, persuadindo seu interlocutor por meio de sua capacidade de interação. TURING, Alan Mathison - *Op. cit.*, pp. 433-434.

[82] A chamada "Conferência de Dartmouth" conduzida por MCCARTY e MINSKY em Hanover, New Hampshire, em 1956 é considerada o berço da inteligência artificial. Foi onde se formulou o seguinte enunciado: "*every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it*". MCCARTHY, John; MINSKY, Marvin; ROCHESTER, Nathan; SHANNON, Claude - *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*.

[83] ANYOHA, Rockwell - *The History of Articial Intelligence: Can Machines Think?*, p. 2.

[84] O primeiro "programa" de inteligência artificial teria sido desenvolvido por ALLEN NEWELL, CLIFF SHAW, e HERBERT SIMON e apresentado na aludida Conferência de Dartmouth. Tratava-se do "*Logic Theorist*", "*designed to mimic the problem solving skills of a human*". ANYOHA, Rockwell - *Can Machines Think?*.

[85] Até a década de 1950, os computadores não dispunham de memória para armazenar dados, apenas executavam comandos e o aluguel mensal de uma máquina chegava a custar $ 200.000,00, de modo que apenas prestigiosas universidades e grandes empresas podiam dispor dos recursos para contar com esses dispositivos. ANYOHA, Rockwell - *Op. cit.*, p. 3.

Em seguida, entre os anos de 1974 e 1980 ocorreu o chamado "primeiro inverno da inteligência artificial", com a derrocada do otimismo que havia surgido no período anterior. Sem o necessário financiamento, os estudos foram suspensos nos institutos mais relevantes[86].

A segunda onda de expansão da IA teve início em 1980, com a retomada dos investimentos e o desenvolvimento de algoritmos mais eficientes, assim como de novas técnicas, como o "aprendizado profundo" ("*deep learning*"), por meio do qual os computadores poderiam ganhar experiência corrigindo erros antes cometidos[87].

Foi o momento em que nasceram os sistemas especialistas ("*expert systems*"). "programas que respondem a perguntas ou resolvem problemas sobre um domínio específico de conhecimento, usando regras lógicas derivadas do conhecimento de especialistas")[88]. Isso se deu graças também ao suporte financeiro japonês especialmente quanto ao programa "*Fifth generation computer project*", o qual contou com um aporte de quase 1 bilhão de dólares[89].

O segundo inverno ("*second AI winter*") descreve o período compreendido entre os anos de 1987 e 1993 e marcou o declínio do entusiasmo com a IA. Embora tenha havido uma desaceleração nos avanços tecnológicos nesse segmento, foram iniciadas novas abordagens (como é o caso da robótica)[90].

A atual fase da IA teria se iniciado em 1993 e indica a retomada do interesse pelo desenvolvimento de sistemas inteligentes, com resultados bastante alvissareiros e impactantes[91], graças particularmente à ampliação exponencial da capacidade de processamentos dos computadores e à introdução de técnicas mais avançadas[92].

---

[86] AGGARWAL, Alok - *The Birth of AI and The First AI Hype Cycle*.
[87] ANYOHA, Rockwell - *Op. cit.*, p. 3.
[88] Seu principal arquiteto foi o cientista da computação EDWARD FEIGENBAUM, para quem: "*Knowledge is power, and the computer is an amplifier of that power. We are now at the dawn of a new computer revolution... Knowledge itself is to become the new wealth of nations*". Fonte: https://computerhistory.org/profile/edward-feigenbaum/ [Consult. 8 jan. 2021].
[89] ANYOHA, Rockwell - *Op. cit.*, p. 3.
[90] ANYOHA, Rockwell - *Op. cit.*, p. 3.
[91] Costumam ser citados como exemplos dos incríveis potenciais da IA as derrotas impostas aos seres humanos por máquinas inteligentes: a) no Xadrez - em maio de 1997, o computador desenvolvido pela IBM Deep Blue ganhou do campeão russo Garry Kasparov; b) no Jeopardy - em fevereiro de 2011, outro computador desenvolvido pela IBM, Watson, derrotou Brad Rutter e Ken Jennings); e c) no Go - em março de 2016, Lee Sedol, mestre do complexo e antigo jogo de tabuleiro chinês, perdeu para o Alpha Zero, sistema de IA criado pela Google (curiosamente, a empresa DeepMind desenvolvedora do sistema havia sido fundada pelos pioneiros da IA e organizadores do Congresso de Dartmouth, tendo sido adquirida pela Google em 2014). Entre outras fontes, pode ser indicada: https://www.theguardian.com/technology/2017/dec/07/alphazero-google-deepmind-ai-beats-champion-program-teaching-itself-to-play-four-hours.
[92] ANYOHA, Rockwell - *Op. cit.*, p. 3.

Em uma síntese desses principais momentos, o grupo empresarial GARTNER decompôs em três etapas a trajetória evolutiva da IA:

Figura 2.

Ciclo de "entusiasmo" e evolução da IA[93]

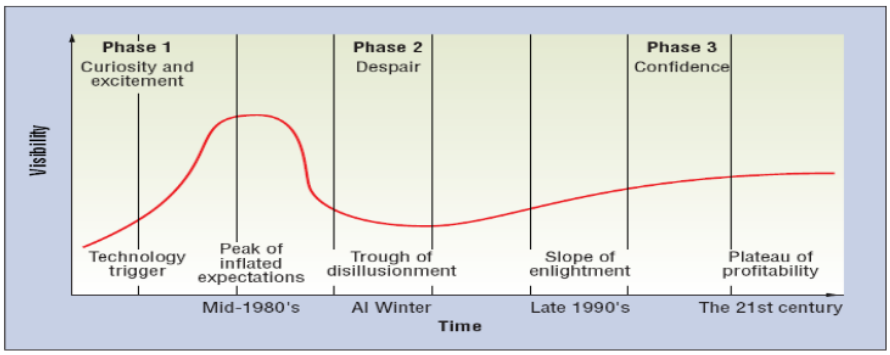


Figure 1. The Hype Cycle and AI winter [Menzies03]

Há alguns teóricos que atribuem ao momento em curso o atingimento de uma etapa denominada "*Artificial Narrow Intelligence*" ("*Inteligência Artificial Estreita*" em uma tradução literal).

Seriam três estágios progressivos: a) "*Artificial Narrow Intelligence*" - aplicada a áreas específicas, sem a capacidade de resolver problemas em outros segmentos de forma autônoma, supera a inteligência humana em seus campos de atuação particulares; b) "*Artificial General Intelligence*" - aplicada a diversas áreas, com a capacidade de resolver problemas em outros segmentos de forma autônoma, supera a inteligência humana nos campos em que atua; e c) "*Artificial Super Intelligence*" - aplicada a qualquer área, com a capacidade de resolver problemas em qualquer segmentos de forma autônoma e instantânea, supera a inteligência humana em todos os campos[94].

1.3. Tendências

A evolução dos sistemas ou agentes inteligentes parece rumar na direção do desenvolvimento das principais características (ou dimensões) da inteligência humana, quais

[93] SMITH, Chris; MCGUIRE, Brian; HUANG, Ting & YANG, Gary – *The History of Artificial Intelligence*, p. 20.
[94] KAPLAN, Andreas & HAENLEIN, Michael - *Op. cit.*, p. 3.

sejam: inteligência cognitiva (*v.g.*, competências relativas ao reconhecimento de padrões e ao pensamento sistemático), inteligência emocional (*v.g.*, adaptabilidade, autoconfiança, autoconsciência emocional e orientação para realização de metas), inteligência social (*v.g.*, empatia, trabalho em equipe e liderança inspiradora) e criatividade artística, única potência antes (e ainda por alguns) tida por inatingível[95]. A mais próxima do atingimento de tal objetivo seria a "inteligência artificial humanizada" ("*humanized AI*"), de acordo com o seguinte quadro comparativo:

Tabela 1.

Quadro comparativo de competências cognitivas [96]

	Expert Systems	Analytical AI	Human-Inspired AI	Humanized AI	Human Beings
Cognitive Intelligence	×	✓	✓	✓	✓
Emotional Intelligence	×	×	✓	✓	✓
Social Intelligence	×	×	×	✓	✓
Artistic Creativity	×	×	×	×	✓
		Supervised Learning, Unsupervised Learning, Reinforcement Learning			

Haveria duas abordagens, em essência, para a implementação de tais sistemas. A primeira poderia ser descrita como "de cima para baixo" ("*top-down approach*", simbólica ou baseada no conhecimento), em que a inteligência seria formalizada por meio de regras ao estilo "se-então" ("*if- then*"), como no caso dos agentes especialistas ("*exper systems*" - v.g. Deep Blue), não qualificados, a rigor, como sistemas inteligentes, por lhes faltar a habilidade de aprendizado a partir de informações externas[97]. Já a segunda, "de baixo para cima" ("*bottom-up approach*", conexionista ou baseada na conduta), simula as condições de

[95] KAPLAN, Andreas & HAENLEIN, Michael - *Op. cit.*, p. 4. Não se pode olvidar a existência de tentativas de produção de conteúdo artístico por meio de sistemas artificialmente inteligentes. Podem ser lembrados três exemplos emblemáticos: o dos Estúdios Botnik que teriam se utilizado de uma ferramenta de IA para escrever um novo capítulo (com o título "*Harry Potter and the Portrait of What Looked Like a Large Pile of Ash*") dentro da saga Harry Potter contendo três páginas inéditas, mas coerentes com o estilo da série (https://botnik.org/studios.html); do projeto "*Next Rembrandt*" que criou uma gravura gera eletronicamente e impressa com realismo (por meio de uma impressora 3D), a partir do estudo das características comuns aos quadros do pintor holandês (https://www.nextrembrandt.com); e da Spotify, que, em parceria com o grupo RZO e mesclando de forma colaborativa a intervenção criativa humana e um sistema de rede neural, analisou composições legadas pelo rapper Sabotage (falecido em 2003) para criar a música "Neural" (https://www.b9.com.br/68359/spotify-lanca-rap-criado-com-inteligencia-artificial/).

[96] KAPLAN, Andreas & HAENLEIN, Michael - *Op. cit.*, p. 5.

[97] *Id., ibid*.

funcionamento do cérebro humano, operando a partir de redes neurais sintéticas no manuseio de dados massivos, com capacidade de aprendizado autônomo (v.g. Alpha Go) [98].

O aprendizado pode se dar de três modos, fundamentalmente: a) com supervisão ("*supervised learning*") - mapeando os dados de entrada ("*input*") e de saída ("*output*"), que seriam "etiquetados" e monitorados; b) sem supervisão ("*unsupervised learning*") - nos quais apenas os dados de entrada ("*input*") seriam mapeados; e c) mediante estímulos (ou reforços positivos - "*reinforcement*") - os dados de saída ("*output*") seriam variáveis e dimensionados de acordo com as decisões tomadas pelo sistema[99].

Algumas das tendências de curto prazo que poderiam ser destacadas, nesse contexto, seriam[100]: i) a robotização com IA incorporada – fruto da integração entre robótica e ciência da computação; ii) a produção automatizada de textos com linguagem semelhante à "humana" – um dos promissores exemplos seria a "*Generative Pre-trained Transformer* 3", um modelo autorregressivo para modelagem de linguagem natural; iii) a transposição da IA para a nuvem – que pode ampliar os recursos computacionais disponíveis para a execução de sistemas artificialmente inteligentes; a incorporação de IA (AIOT) na internet das coisas (IoT) – com a possibilidade de ampliar, exponencialmente, a rede "sensorial" dos sistemas de IA; iv) a implementação das "redes adversariais generativas" ("*Generative Adversarial Networks – GANs*") – visto como o futuro do aprendizado profundo ("*deep learning"*), capaz de criar imagens inéditas e realistas.

Para o longo prazo, em que os cenários revelam-se marcadamente especulativos, há um percepção predominantemente otimista em torno da concretização de potencialidades disruptivas da IA em uma miríade de segmentos, entre os quais se poderia mencionar: i) o transporte – com a implantação do último estágio de automação dos veículos plenamente autônomos (para alguns, isso estaria situado em uma perspectiva de médio ou mesmo curto prazo); ii) saúde – na medicina diagnóstica de altíssima profundidade analítica, desenvolvimento de remédios altamente específicos e personalizados, além das intervenções cirúrgicas realizadas por robôs dotados de grande capacidade cognitiva e motora; iii) integração ("*merging*") entre inteligência humana (orgânica) e inteligência artificial ("*transhumanism*")[101].

---

[98] *Id., ibid*.
[99] KAPLAN, Andreas & HAENLEIN, Michael - *Op. cit.*, p. 6.
[100] Fonte: https://towardsdatascience.com/top-5-ai-trends-that-will-change-the-landscape-of-the-future-bd398ff9db55 [Acesso em 03/03/21]
[101] FAST, Ethan & HORVITZ, Eric - *Long-Term Trends in the Public Perception of Artificial Intelligence*, p. 2.

Ainda no horizonte de longo prazo, identificam-se algumas importantes preocupações, tais como: a perda de controle sobre sistemas de IA mais sofisticados, a perda de postos de trabalho, o uso militar de semelhantes tecnologias, a ausência de parâmetros éticos definidos para guiar o desenvolvimento e implementação de tais ferramentas[102].

Um dos pontos altos das previsões sobre o futuro, muito comumemente explorados em obras de ficção científica, a singularidade ("*singularity*")[103] pode ser vista: i) tanto sob a ótica positiva, enquanto forma de alcançar o shangrilá da imortalidade; ii) como negativa, desidratação do valor intrínseco da vida humana, ante a possibilidade de transferência da consciência para a nuvem[104].

A presente investigação concentra-se, contudo, nos impactos da IA em uma perspectiva de curto e médio prazo, o que exige uma análise das suas principais características, associadas aos seus riscos mais relevantes.

1.4. Características

1.4.1. Opacidade ("*black box-effect*")

A opacidade epistêmica dos mecanismos artificialmente inteligentes tem sido vista como um óbice à sua confiabilidade. A demanda por transparência nos processos que levam aos resultados alcançados pela máquina, além de representar uma demanda "emocional"[105], por assim dizer, atende às exigências de sindicabilidade ou "*accountability*", a permitir o seu controle judicial.

Em princípio, a maior parte dos agentes de IA opera com análises correlacionais fundadas em estatísticas e não com a investigação de cadeias causais. Portanto, são suscetíveis a alguma falibilidade nas conclusões que podem alcançar. A ausência de controle ou supervisão dos parâmetros utilizados mostra-se especialmente grave, nessa perspectiva.

---

[102] FAST, Ethan & HORVITZ, Eric – *Op.cit.*, p. 2.

[103] Trata-se de expressão matemática empregada em diversos contextos diferentes. No domínio da inteligência artificial, o termo designa essencialmente o momento em que os sistemas de IA serão capazes de ingressar em uma espiral ascendente de auto-desenvolvimento de modo a superar a inteligência humana e tomar as rédeas do processo de desenvolvimento de outros agentes artificialmente inteligentes. LADER, Bhagirath Kumar - *Some Thoughts on Artificial Intelligence Singularity*, p. 1.

[104] FAST, Ethan & HORVITZ, Eric – *Op.cit.*, p. 2.

[105] A esse propósito, PARASURAMAN, ZEITHAML e BERRY oferecem uma ótima ilustração: "*While AI systems are already capable of diagnosing X-ray images as proficiently as, or even better than, most physicians, most consumers would have a difficult time trusting the verdict of a machine. [...] A necessary precondition to this confidence, besides better explanation, is that firms do not overpromise on the potential of AI.*". *Apud* KAPLAN, Andreas & HAENLEIN, Michael - *Op. cit.*, p. 8.

O caso *State v. Loomis* afigura-se emblemático nesse ponto. "*Correctional Offender Management Profiling for Alternative Sanctions*" (COMPAS) consiste em uma ferramenta de gerenciamento de casos ("*case management*") e suporte à decisão desenvolvida pela empresa Northpointe, usada por tribunais norte-americanos para avaliar a probabilidade de um réu se tornar reincidente. Entre outras utilidades, permite avaliar a possibilidade de liberação do acusado mediante fiança, além de oferecer contribuições para definição da pena e indicar a forma mais adequada de supervisão da sua execução[106].

Em fevereiro de 2013, ERIC LOOMIS foi condenado a seis anos de prisão por ter se evadido do local (desrespeitando as ordens dirigidas por autoridades policiais) em que teria ocorrido um tiroteio na cidade de Lacrosse (Wisconsin). Ao fixar a pena, a sentença valeu-se do auxílio do aludido sistema[107].

Ao recorrer à Suprema Corte de Wisconsin LOOMIS afirmou que o uso de uma avaliação de risco produzido por um programa teria violado o seu direito ao devido processo por três razões, em essência: a) ofenderia seu direito de ser sentenciado com base em informações precisas, em parte porque a natureza proprietária do sistema o impediria de ter acesso aos dados utilizados (e mesmo ao algorítimo) para avaliar sua precisão; b) malferiria a garantia de individualização da pena; e c) teria incorrido em conduta discriminatória ao se valer de fundamentos atuariais relativos ao gênero[108].

A Suprema Corte de Wisconsin não acolheu o apelo do réu, refutando todas as suas alegações aos seguintes fundamentos: a) a decisão não teria se baseado exclusivamente nas conclusões apresentadas no relatório do COMPAS; b) os parâmetros gerais adotados pelo sistema para seu prognóstico atuarial teriam sido publicizados por meio de manual divulgado pela própria empresa desenvolvedora (*Practitioner's Guide to COMPAS Core*); c) teriam sido indicados diversos elementos no laudo gerado pelo sistema que permitiriam não apenas a individualização a análise de risco de reincidência como a sua impugnação; e d) dadas as estatísticas indicadoras de maior nível de reincidência entre os homens, não teria ocorrido discriminação por parte do programa de suporte à decisão, mas antes teria sido garantida uma maior acurácia na fixação da pena[109].

---

[106] NORTHPOINTE - *Practitioner's Guide to COMPAS Core*, p. 1.

[107] Entre outras fontes, a síntese do caso encontra-se em matéria publicada na Harvard Law Review sob o título "*State v. Loomis - Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing*", em 10/03/17. Disponível em: https://harvardlawreview.org/2017/03/state-v-loomis/

[108] HARVARD LAW REVIEW - "*State v. Loomis - Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing*".

[109] O inteiro teor da decisão encontra-se em https://www.wicourts.gov/sc/opinion/DisplayDocument.pdf?content=pdf&seqNo=171690

Não obstante, o referido Tribunal destacou que o sistema não poderia ser utilizado para determinar o tipo de pena, assim como impôs a exigência de divulgação aos juízes adeptos do COMPAS dos seguintes alertas: "[...] primeiro, a natureza proprietária do COMPAS impede a divulgação de como as pontuações de risco são calculadas; segundo, os escores do COMPAS são incapazes de identificar indivíduos específicos de alto risco, porque esses escores dependem de dados de grupo; terceiro, embora o COMPAS dependa de uma amostra nacional de dados, não houve qualquer estudo de validação cruzada para a população de Wisconsin; quarto, estudos levantaram questões sobre se os escores do COMPAS desproporcionalmente classificam os ofensores pertencentes a minorias como tendo maior risco de reincidência; e quinto, o COMPAS foi desenvolvido especificamente para auxiliar o Departamento de Correções na tomada de determinações pós-sentença"[110].

Houve recurso à Suprema Corte dos EUA, mas a petição de "*writ of certiorari*" foi rejeitada em 26 de junho de 2017[111].

O principal desafio aqui decorreria da necessidade de conciliação do respeito aos interesses privados (ligados a direitos como propriedade privada e intelectual, assim como livre iniciativa) consistentes na proteção dos dados sensíveis relativos ao algoritmo desenvolvido pela empresa privada, de um lado, e a preservação do direito de acesso a informações (vinculado sobretudo a garantias de ampla defesa, contraditório e isonomia) que permitam auditar as operações desenvolvidas pelo sistema e suas respectivas conclusões, de outro.

Nessa esteira, pode ser mencionado o princípio da transparência judicial ("*judicial transparency*") formulado pelos integrantes do Instituto "Future for Life" (FLI)[112], segundo o qual: "qualquer envolvimento de um sistema autônomo na tomada de decisões judiciais deve fornecer uma explicação satisfatória passível de auditoria por uma autoridade humana competente"[113].

---

[110]HARVARD LAW REVIEW - *State v. Loomis - Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing*.

[111] Andamento disponível em https://www.supremecourt.gov/docketfiles/16-6387.htm

[112] Eis a missão do Instituto: "*To catalyze and support research and initiatives for safeguarding life and developing optimistic visions of the future, including positive ways for humanity to steer its own course considering new technologies and challenges*". Disponível em https://futureoflife.org/team/ [Consult. 29 dez. 2018]

[113] No original: "*any involvement by an autonomous system in judicial decision-making should provide a satisfactory explanation auditable by a competent human authority*". O princípio faz parte do conjunto conhecido como "Princípios de IA de Asilomar" ("Asilomar AI Principles"), formulados em 2017 em evento organizado pelo FLI e subscritos por milhares de cientistas, pensadores e líderes de tecnologia, tai como STEPHEN HAWKING, ELON MUSK, JAAN TALLIN, DEMIS HASSABIS, entre outros. Disponível em https://futureoflife.org/ai-principles/ [Consult. 29 dez. 2018]

Não pode deixar de ser mencionado, outrossim, o princípio da transparência, imparcialidade e equidade enunciado pela já mencionada CEPEJ do Conselho da Europa[114].

De acordo com o supramencionado princípio, deve-se procurar alcançar um equilíbrio entre: a) "propriedade intelectual de certos métodos de processamento" e; b) "necessidade de transparência (acesso ao processo de design), imparcialidade (ausência de parcialidade), justiça e integridade intelectual (priorizando os interesses da justiça)". Tal preocupação deve perpassar toda a cadeia de desenvolvimento do projeto (desde a concepção até a execução operacional), pois o processo de seleção e a qualidade e organização dos dados influenciariam diretamente a fase de aprendizado da máquina[115].

Entre as opções para concretizar a medida, a Comissão indica a completa transparência técnica, v.g. por meio de códigos abertos ("*open source code*") e documentação pública, com as naturais restrições decorrentes dos segredos empresariais. Além disso, ressalta que as ferramentas inteligentes poderiam fazer uso de "traduções" para a linguagem natural ("humana") para descrever como os resultados da análises foram obtidos[116]. Por fim, autoridades independentes poderiam auditar e certificar os métodos de processamento dos dados[117].

Outro desdobramento essencial consiste na exigência de expressa manifestação de consentimento como condição de validade para a condução de operações de retenção e manipulação de dados pessoais[118].

Referidas diretrizes mostram-se naturalmente aplicáveis não apenas aos instrumentos adotados pelo Pode Judiciário mas a toda e qualquer forma de tecno-regulação.

No concernente às ferramentas tecno-regulatórias, por vezes, a própria norma encontra-se oculta, especialmente quando se trata de dispositivos computacionais[119] e os reguladores são entes privados (não estatais).

Constata-se, por conseguinte, a possibilidade de se verificar um forte déficit democrático no processo de construção e implementação de mecanismos tecno-regulatórios.

---

[114] CONSELHO DA EUROPA - Carta Europeia de Ética no uso da inteligência artificial em sistemas judiciais e seu ambiente.

[115] CONSELHO DA EUROPA - *Op. cit.*, p. 9.

[116] Na mesma direção, há uma nova linha de pesquisa conhecida como "IA explicável" ("*explainable AI*") focada na extração de regras "ex-post", para que se possa compreender, ainda que de foram rudimentar as operações desenvolvidas pelo sistema. KAPLAN, Andreas & HAENLEIN, Michael - *Op. cit.*, p. 8.

[117] *Id. ibid.*

[118] O regulamento europeu disciplina a questão do consentimento em seus arts.ºs 7.º e 8º. No Direito brasileiro, a diretriz encontra-se positivada no art.º 7.º, I, da Lei 13.709/18.

[119] Nesse sentido, assevera LEENES que: "*When norms are embedded into computer applications or devices, they are not always clear. Sometimes the existence of a norm may even be unknown to the user*".

Daí a necessidade de transpor para tais instâncias de regulação toda a principiologia suprareferida.

Como se vê, o debate em torno da necessidade de "explicabilidade" das atividades cognitivas de tais sistemas emerge de preocupações jurídicas razoáveis e relevantes.

Haveria duas tendências antagônicas no tocante ao enfrentamento do hermetismo algorítmico.

A primeira propõe, por assim dizer, a tradução da "linguagem" das máquinas para uma expressão inteligível aos homens. Trata-se de uma abordagem mais conhecida como "IA explicável" ("*explainable AI*" ou "*XAI*"), que consistiria, em verdade, na definição de regras "*ex-post*" para interpretar as operações desenvolvidas pelo sistema[120].

A esse propósito, cumpre ressaltar que a interpretabilidade apresenta-se como um conceito relevante e correlato ao da explicabilidade. Enquanto o segundo concerne à tentativa de tornar os mecanismos internos da "*machine*" ou "*deep learning*" explicáveis em "termos humanos", o primeiro diz respeito à identificação de relações causais dentro dos processos cognitivos artificiais, i.e., dos nexos existentes entre os dados coletados e as conclusões alcançadas[121].

Há quem proponha, ainda, como possíveis soluções para enfrentar a opacidade, conciliando propriedade privada e interesses individuais: a) o acesso aos dados pessoais mediante assinatura de termo de confidencialidade quando a informações técnicas sensíveis; ou b) a abertura dos sistemas a corpos neutros de especialistas que os auditariam para avaliar sua correção e virtual ofensa a direitos fundamentais[122].

Já a segunda, defendida por autores como BAROCAS e SELBST, defende que a transparência não seria um fim em si mesmo e o ônus resultante da interpretabilidade dos

---

[120] KAPLAN, Andreas & HAENLEIN, Michael - *Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence*, p. 8.

[121] "*Interpretability is about the extent to which a cause and effect can be observed within a system. Or, to put it another way, it is the extent to which you are able to predict what is going to happen, given a change in input or algorithmic parameters. It's being able to look at an algorithm and go yep, I can see what's happening here. Explainability, meanwhile, is the extent to which the internal mechanics of a machine or deep learning system can be explained in human terms. It's easy to miss the subtle difference with interpretability, but consider it like this: interpretability is about being able to discern the mechanics without necessarily knowing why. Explainability is being able to quite literally explain what is happening*." GALL, Richard – *Machine Learning Explainability vs Interpretability: Two concepts that could help restore trust in AI*.

[122] "*One possibility is that in any individual adjudication, the technical aspects of the system could be covered by a protected order requiring their confidentiality. Another possibility is to limit disclosure of the scoring system to trusted neutral experts. Those experts could be entrusted to assess the inferences and correlations contained in the audit trails*". CITRON, Danielle Keats & PASQUALE, Frank - *The Scored Society: Due Process for Automated Predictions*, p. 28

algoritmos repousaria na perda de sua acurácia[123]. A demanda por "*accountability*" exigiria algum nível de simplificação dos modelos, o que implicaria a degradação de sua performance[124].

Como se vê, a superação da opacidade representa um desafio mais complexo do que costuma sugerir o senso comum, o que não exime os desenvolvedores de tais sistemas da responsabilidade quanto às falhas resultantes da inacessibilidade aos seus fluxos operacionais por parte de instâncias de auditagem externa.

1.4.2. Estrutura correlacional e não causal

Entre as soluções de IA atuais mais comumente difundidas, presentemente, os sistemas especialistas atuam, as mais das vezes, a partir da análise correlacional de dados[125].

Vale notar que a maior parte dos sistemas que integram o que se conhece como Inteligência Artificial, hodiernamente, confunde-se com o que os cientistas da computação qualificavam como "Aprendizado de Máquina" ("*Machine Learning - ML*") nas últimas décadas. "ML" resulta de um conjunto de contribuições da Estatística, Matemática, Ciência

---

[123] No mesmo sentido segue MARCELLO PELILLO: "*L'opacità degli algoritmi di machine learning, […], va di pari passo con la loro precisione. Più l'intelligenza della macchina riesce ad estrarre da una mole enorme di dati quelli che le sono utili e apprendere come migliorare la propria performance, più questa operazione sarà oscura per gli utenti, ma anche per gli esperti che volessero correggerla*". Fonte: https://www.unive.it/pag/14024/?tx_news_pi1%5Bnews%5D=6293&cHash=0fada200b6d0fe0ad24d3f18956939ec [Consult. 15 mar. 2021] .

[124] Ponderam SELBST e BAROCAS a esse respeito "*From scholars seeking to "unlock the black box" to regulations requiring "meaningful information about the logic" of automated decisions, recent discussions of machine learning—and algorithms more generally—have turned toward a call for explanation. Champions of explanation charge that algorithms must reveal their basis for decision-making and account for their determinations. But by focusing on explanation as an end in itself, rather than a means to a particular end, critics risk demanding the wrong thing. What one wants to understand when dealing with algorithms—and why one would want to understand it—can vary widely. Often, the goals that motivate calls for explanation could be better served by other means. Worse, ensuring that algorithms can be explained in ways that are understandable to humans may come at the cost of another value that scholars, regulators, and critics hold dear: accuracy. Reducing the complexity of a model, for example, may render it more readily intelligible, but also hamper its performance. For explanation to serve its intended purpose and to find its appropriate place among a number of competing values, its champions need to consider what they hope it to achieve and what explanations actually offer [...] machine learning may produce models that differ so radically from the way humans make decisions that they resist sense-making*". BAROCAS, Solon and SELBST, Andrew D. - *Regulating Inscrutable Systems*, p. 1.

[125] Exemplo curioso de conclusões apoiadas em estudos meramente correlacionais pode ser encontrado em https://www.businessinsider.com/chocolate-consumption-vs-nobel-prizes-2014-4. A matéria aborda a correlação estabelecida entre o nível de consumo de chocolate por diversos países e a conquista do Prêmio Nobel, em pesquisa publicada no *New England Journal Of Medicine*. Segundo o condutor da investigação, FRANZ MESSERLI: "*Since chocolate consumption could hypothetically improve cognitive function not only in individuals but also in whole populations, I wondered whether there would be a correlation between a country's level of chocolate consumption and its population's cognitive function. To my knowledge, no data on overall national cognitive function are publicly available. Conceivably, however, the total number of Nobel laureates per capita could serve as a surrogate end point reflecting the proportion with superior cognitive function and thereby give us some measure of the overall cognitive function of a given country*".

computacional e de outras disciplinas conexas para projetar algoritmos que processam dados, fazem previsões e ajudam a tomar decisões[126].

Trata-se de uma perspectiva fundamentalmente correlacional, sem qualquer avaliação de eventuais cadeias causais.

Correlação poderia ser definida como o grau em que dois ou mais atributos ou medidas no mesmo grupo de elementos mostram uma tendência para variar juntos[127]. Duas variáveis seriam correlacionadas se suas subestruturas subjacentes estivessem conectadas[128].

O recurso à via correlacional produz riscos epistêmicos sensíveis. Alguns matemáticos vêm alertando para a circunstância de que as vulnerabilidades ampliam-se com o alargamento do volume de dados manuseados pelos sistemas ("*big data*")[129].

Uma das principais fontes de equívoco consiste na identificação entre correlação e causalidade por parte de alguns dos que desenvolvem ou fazem uso de sistemas de IA para extrair conclusões e tomar decisões.

Ora, como se sabe, "correlação não implica causalidade", tal como lembravam os filósofos clássicos ao repelir a falácia "*cum hoc ergo propter hoc*" ("com isto, logo por causa disto")[130]. Não se revela logicamente viável inferir nexo causal de fenômenos relacionados entre si. A causa apresenta-se como uma possibilidade e não como necessidade lógica, nesse contexto.

Entre os que se aprofundaram no conceito de causalidade, HUME legou diversas definições e princípios[131] relevantes para uma compreensão mais plena do termo em sua

---

[126] JORDAN, Michael - *Artificial Intelligence—The Revolution Hasn't Happened Yet*, p. 3.
[127] MARWALA, Tshilidzi - *Causality, Correlation and Artificial Intelligence for Rational Decision Making*, p. 2.
[128] MARWALA, Tshilidzi - *Op. Cit.*, p. 2.
[129] "*The mathematicians Cristian Sorin Calude and Giuseppe Longo point to the risk of a deluge of false correlations in big data: the larger a database used for correlations, the greater the chances of finding recurrent patterns and the greater the chances of making errors. 41 What may appear as regularities for an AI (recurrent links between different data, concepts, contexts or lexical groups) may actually be random. Even if the argument of the two mathematicians should not be generalised too hastily, they note that in certain vast sets of numbers, points or objects, regular random patterns appear and it seems impossible to distinguish them algorithmically from patterns revealing causalities*". RONSIN, Xavier - *In-depth study on the use of AI in judicial systems, notably AI applications processing judicial decisions and data*, p. 35.
[130] Entre outros, a ideia pode ser vista explorada de forma peculiar em TUFTE, Edward R - *The Cognitive Style of PowerPoint: Pitching Out Corrupts Within*.
[131] DAVID HUME enunciou oito princípios da causalidade: "*1. The cause and effect are connected in space and time. 2. The cause must happen before the effect. 3. There should be a continuous connectivity between the cause and the effect. 4. The specific cause must at all times give the identical effect, and the same effect should not be obtained from any other event but from the same cause. 5. Where a number of different object give the same effect, it ought to be because of some quality which is the same amongst them. 6. The difference in the effects of two similar events must come from that which they are different. 7. When an object changes its dimensions with the change of its cause, it is a compounded effect from the combination of a number of different effects, which originate from a number of different parts of the cause. 8. An object which occurs for any time in its full exactness without*

conotação filosófica. Segundo o filósofo escocês, causa poderia ser concebida como "um objeto precedente e contíguo a outro, e onde todos os objetos que se assemelham ao primeiro são colocados em relações de precedência e contiguidade semelhantes àqueles objetos que se assemelham ao último"[132].

Sobreleva notar que o ceticismo do pensador nesse particular levou-o a sustentar a carência de recursos cognitivos na mente humana para observar as relações causais com alguma acurácia [133].

A rigor, ambas as perspectivas (correlacional e causal) seriam relevantes para qualquer processo decisório racional[134]. No âmbito jurisdicional, por óbvio, a dimensão causal garantiria maior confiabilidade epistêmica na subsunção de condutas a normas jurídicas, sobretudo quando se trata da imputação de responsabilidade nos âmbitos civil e criminal. Contudo, a avaliação correlacional pode suprir eventuais claros no processo de reconstrução dos supedâneos fáticos para aplicação das regras de Direito.

Nessa esteira, cumpre observar que dotar agentes artificialmente inteligentes da capacidade de realização de inferência causal tem sido vista como um desafio apto a ser superado por novos modelos de IA. Em tal perspectiva, entre as diversas tentativas conduzidas para definir uma infraestrutura matemática consistente com o raciocínio causal, poderiam ser mencionados: a) o Modelo Estrutural Causal ("*Structural Causal Models - SCM*") - desenvolvido por JUDEA PEARL, possibilitaria a aferição causal a partir de "Gráficos Acíclicos Direcionados" ("*Directed Acyclic Graphs - DAG*") e mediante a aplicação da matemática da intervenção[135]; e b) o Modelo Causal de Rubin ("*Rubin Causal*

---

any effect, is not the only cause of that effect, but needs to be aided by some other norm which may advance its effect and action". HUME, David - *A Treatise of Human Nature*, p. 95.

[132] No original: "*an object precedent and contiguous to another, and where all the objects resembling the former are placed in like relations of precedency and contiguity to those objects that resemble the latter*". HUME, David - *Op. cit.*, p. 94.

[133] Pode-se afirmar, por conseguinte, que à luz das contribuições de HUME não haveria diferenças substanciais entre o intelecto humano e os atuais sistemas de inteligências artificial nesse ponto, uma vez que, de certo modo, ambos partiriam de inferências de natureza predominante correlacional e não propriamente causal. Eis excerto que, em certa medida, corrobora tal conclusão: "*According to the precedent doctrine, there are no objects, which by the mere survey, without consulting experience, we can determine to be the causes of any other; and no objects, which we can certainly determine in the same manner not to be the causes. Any thing may produce any thing*". HUME, David - *Op. cit.*, p. 94.

[134] "[...] *the basic elements of rational decision making are causality and correlation. Correlation is important because it fills in the gaps that exist because of the limitation of information in decision making. Furthermore, causality is important in decision making because it is through causal loops (if this then act so) that an appropriate cause of action is chosen from many possible causes of actions and thus reducing the complexity of the problem*". MARWALA, Tshilidzi - *Op. Cit.*, p. 2.

[135] "*The mathematics of intervention, a necessary component for inferring causality, are provided in SCMs through the "do-operator." When assessing whether X is causing Y, one learns, through this framework, that in general, Pr(Y=y|X=x)≠ Pr(Y=y|do(X=x)), where do(X) denotes an intervention on the random variable X and Pr(.) denotes*

Model - RCM") - também conhecido como "Estrutura de Resultados Potenciais", foi formulado por DONALD RUBIN e confronta os resultados potenciais (Y) de um experimento intervencionista em que um ponto de dados (v.g. um indivíduo) é exposto a diferentes níveis de X[136].

Ainda se mostram incipientes tais linhas de investigação, mas revelam potencialidades promissoras de aprimoramento da eficiência epistêmica dos agentes artificialmente inteligentes, a partir da mitigação ou mesmo superação dos modelos correlacionais.

1.4.3. Complexidade e confiabilidade

Sistemas artificialmente inteligente revelam-se, intrinsecamente, complexos, embora sua complexidade não seja linear[137].

De sua complexidade deriva o risco de falibilidade, cuja contenção desafia técnicos, cientistas e teóricos, ante os desafios resultantes, dentre outros fatores, do design técnico do sistema, do ambiente exterior e da interação com os usuários[138].

A expressão "confiabilidade" ("*reliability*") descreve as preocupações em torno das falhas operacionais e sistêmicas que podem resultar da implementação de soluções de IA e se mostram especialmente preocupantes em situações nas quais o direito à vida pode ser

the probability of a random variable (note that here, I only considered whether X is causing Y, such relationship is asymmetric in a causal sense and hence, whether Y causes X needs a separate causal analysis). The left-hand side of the formula above points to conditional probabilities (not causal in general), whereas elucidation of a causal relationship between variables requires incorporation of intervention (right-hand side of the formula) and using mathematics of the do-calculus". KHADEMI, Aria - *Causal Reasoning: A Fairly Overlooked Piece in Artificial Intelligence*.

[136] "*In order to elucidate whether X causes Y, this framework, speaking at a high level, contrasts the potential outcomes (Y) of an interventional experiment in which a data point (say, an individual) is exposed to different levels of X. For example, if we are to examine the causal effect of Advil on headache, we would contrast the outcomes (whether one has a headache) of an interventional experiment treating (taking Advil) and controlling (not taking Advil). The gold standard experiment for finding such contrast of outcomes is a Randomized Controlled Trial (RCT), but RCTs are not always feasible. The details of experimental design as well as those of how to infer causality in the absence of RCTs is out of the scope of this article, but I would be more than happy to discuss further with interested readers*". KHADEMI, Aria - *Op. cit*.

[137] GWERN - *Complexity no Bar to AI*.

[138] *International Committee of the Red Cross - Autonomy, artificial intelligence and robotics: Technical aspects of human control,* p. 2.

comprometido, como é o caso do uso da tecnologia para fins militares (sobretudo de armas letais)[139] ou no campo da medicina[140].

Segurança, em tais contextos, revela-se essencial, o que pode ser obtido a partir da construção de modelos algorítmicos robustos e resilientes. Para avaliar sua confiabilidade, têm sido desenvolvidas métricas que permitam a mensuração objetiva de aludidos atributos[141], além de serem considerados sistemas de IA destinados à realização dessa avaliação[142].

O tema tem sido objeto de diversos marcos regulatórios e encontra-se presente também na proposta de regulação europeia[143]. Trata-se, aliás, do epicentro das preocupações que levaram a Comissão Europeia a propor novas regras destinadas a garantir a confiabilidade dos sistemas de IA[144].

Embora a proposta regulatória venha a ser analisada em capítulo próprio da presente dissertação, vale realçar que, no caso dos denominados "sistemas de alto risco"[145], entre as exigências impostas para assegurar a contenção das vulnerabilidades operacionais, podem ser mencionadas: i) a implementação de sistema de gestão de risco; ii) a manutenção de programa de governança de dados; iii) o rigoroso registro dos eventos ("log") em todas as fases da sua operação; e iv) a existência de supervisão humana[146].

---

[139] *"Predictability and reliability are at the heart of discussions about autonomy in weapon systems, since they are essential to achieving compliance with international humanitarian law and avoiding adverse consequences for civilians. They are also essential for military command and control. It is important to distinguish between: reliability – a measure of how often a system fails; and predictability – a measure of how the system will perform in a particular circumstance. Reliability is a concern in all types of complex system, whereas predictability is a particular problem with autonomous systems. There is a further distinction between predictability in a narrow sense of knowing the process by which the system functions and carries out a task, and predictability in a broad sense of knowing the outcome that will result*". International Committee of the Red Cross – *Op. cit., p. 2*.

[140] A esse respeito, entre outros, pode ser citado o seguinte estudo: BALAGURUNATHAN, Yoganand & MITCHELL, Ross & EL NAQA, Issam - *Requirements and reliability of AI in the medical context*.

[141] É o que se pretende propor, ilustrativamente, nesse "paper": JHA, Susmit & RAJ, Sunny & FERNANDES, Steven - *Attribution-Based Confidence Metric For Deep Neural Networks*.

[142] Na seara militar, há notícia nesse sentido em matéria disponível no portal Tech Xplore, sob o título: "*Researchers measure reliability, confidence for next-gen AI*". Disponível em: https://techxplore.com/news/2020-03-reliability-confidence-next-gen-ai.html [Acesso em 22 jun. 2021].

[143] O texto completo pode ser encontrado em: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence [Acesso em 21 abr. 2021]

[144] O anúncio foi feito em 21 de abril do ano em curso (2021). Fonte: https://ec.europa.eu/commission/presscorner/detail/pt/ip_21_1682 [Acesso em 21 abr. 2021].

[145] Os quais se referem a sistemas ligados a infraestruturas críticas, segurança de produtos, aplicação coercitiva da lei e, mesmo, "aplicação coercitiva de lei", dentre outros (art.º 6.º da proposta de regulação).

[146] Fonte: https://ec.europa.eu/commission/presscorner/detail/pt/ip_21_1682 [Acesso em 21 abr. 2021]

1.4.4. Autonomia e Imprevisibilidade

Embora o atual estágio de desenvolvimento dos sistemas de IA não pareça evidenciar a viabilidade de existência de agentes artificialmente inteligentes dotados de plena autonomia, divisa-se no horizonte a superveniência de tecnologias que talvez venham a permitir a criação de entes tão autônomos quanto os seres humanos.

Seja como for, pode-se afirmar que, em certa medida, a imprevisibilidade da conduta de tais agentes é diretamente proporcional ao nível de autonomia de que dispõe[147].

A partir de uma didática e simplificada explicação, a autonomia de um sistema ou função pode ser descrita como um "circuito fechado" (ou "*closed loop*"), composto de três estágios que fazem lembra o axioma aristotélico[148]: i) sentir ("*sense*") – captura de informações do ambiente exterior, por meio de sensores; ii) pensar ("think") – processamento dos dados recebidos, mediante o uso de algoritmos; e iii) agir ("*act*") – realização de intervenção no mundo exterior, a partir da análise realizada[149].

A imprevisibilidade representa, nesse contexto, uma dificuldade em se antever sobretudo como se desenvolverá esse terceiro momento, mas também remetem à necessidade de se avaliar (ou mesmo "auditar") as duas fases anteriores, o que se revela especialmente árduo, considerada a já abordada opacidade inerente a boa parte das abordagens da IA.

Entre os riscos decorrentes da autonomia, encontra-se a dificuldade de imputação de responsabilidade aos desenvolvedores, proprietários ou usuários de tais sistemas. Um dos exemplos mais emblemáticos refere-se aos acidentes provocados por veículos autônomos[150].

Algumas respostas jurídicas a esse desafiador problema tem sido formuladas mais recentemente, conforme se passa a expor.

---

[147] A esse respeito, eis o que lembra relatório da Cruz Vermelha sobre o assunto: "In a broad sense, all autonomous systems are unpredictable to a degree because they are triggered by their environment. However, developments in the complexity of software control systems – especially those based on AI and machine learning – add unpredictability in the narrow sense that the process by which the system functions is unpredictable." International Committee of the Red Cross – *Op. cit., p. 2*.

[148] "*Nihil est in intellectu quod non prius fuerit in sensu*". Em tradução livre: "nada está na mente que não tenha sido obtido a partir dos sentidos". Fonte: https://academic.oup.com/jhmas/article-abstract/XXV/1/77/737664?redirectedFrom=fulltext [Acesso em 23/02/21]

[149] International Committee of the Red Cross – *Op. cit., p. 7*.

[150] A esse respeito, vide GRIEMAN, Keri - *Hard Drive Crash An Examination of Liability for Self-Driving Vehicles*.

1.5. Natureza Jurídica dos Agentes Artificialmente Inteligentes

Há um crescente interesse doutrinário sobre uma questão que até bem pouco tempo repousava apenas nos livros de ficção científica. Entre os juristas, a matéria sempre soou meramente especulativa e própria a nefelibatas e futurólogos desprovidos de qualquer sentido prático[151].

A natureza jurídica dos agentes eletrônicos artificialmente inteligentes[152] passou, contudo, a frequentar ponderáveis obras de sérios civilistas agora preocupados com as implicações jurídicas das decisões tomadas por tais "seres" atualmente privados de qualquer traço de personalidade jurídica.

Cumpre ter em conta não se tratar de meros instrumentos para a implementação de transações jurídicas por meio eletrônico, tal como uma "máquina de refrigerante"[153]. Aqui não se está referindo a meros programas de computador que realizam tarefas limitadas e previsíveis. Seus sofisticados algoritmos ensejam aprendizado, modificação de conduta em resposta a mudanças no ambiente externo, de forma a permitir a tomada de decisões, sem qualquer supervisão "humana"[154].

Uma das principais angústias a afligirem os estudiosos refere-se à imputação de responsabilidade civil por eventuais danos causados por aludidos agentes, que passaram a tomar decisões com autonomia, graças a suas elevadas capacidades cognitivas[155].

Nesse cenário, brotaram em profusão diversas abordagens teóricas com o objetivo de definir a melhor fórmula para equacionar os desafios jurídicos relativos aos direitos e obrigações decorrentes da atuação de agentes eletrônicos.

---

[151] A discussão não é recente, consoante se depreende do estudo conduzido por LAWRENCE SOLUM no início da década de 1990. SOLUM, Lawrence B. - *Legal Personhood for Artificial Intelligences*.

[152] Cuida-se de subespécie de agentes eletrônicos, os quais na definição de BELLIA, seriam "*a software thing programmed to execute sophisticated commands of a human user*" (BELLIA, Anthony J. – *Contracting with Electronic Agents*, p. 1.051). Já os agentes artificialmente inteligentes consistiriam em avançados programas de computador dotados de capacidades cognitivas complexas baseados em sofisticados algoritmos que permitem a execução de tarefas variadas.

[153] KERR, Ian - *The Arrival of Artificial Intelligence and "The Death of Contract"*, p. 4.

[154] ANDRADE, Francisco & NOVAIS, Paulo & NEVES, José - *Issues on Intelligent Electronic Agents and Legal Relations*, p. 3.

[155] Diversos pensadores entendem que o mais adequado seria manter o centro de imputação de responsabilidade nas pessoas individuais e coletivas envolvidas. É o caso de JOHNSON e VERDICCHIO, os quais enunciam o conceito de agência triádica ("*triadic agency*") a partir da premissa segundo a qual a atribuição de intencionalidade não poderia alcançar entidades algorítmicas. Propõem que haveria, invariavelmente, três elementos: a) o usuário ("*the user*") – que define a meta e delega ao designer a tarefa de conceber o caminho para o seu atingimento; b) o designer ("*the designer*") – que cria o artefato com o objetivo de atingir a meta definida; e c) o artefato ("*the artifact*") – provido de eficácia causal para alcançar a meta proposta. De acordo com os autores, apenas o primeiro e o segundo (se forem pessoas no sentido jurídico) é que poderiam vir a responder juridicamente por eventuais danos causados pelo artefato. JOHNSON, Deborah G. & VERDICCHIO, Mario - *AI, agency and responsibility: the VW fraud case and beyond*.

No espectro heterogêneo de concepções a respeito do assunto, podem ser agrupadas as principais correntes doutrinárias em três blocos, fundamentalmente: a) o dos que se voltam para o passado à busca de uma solução para o problema; b) o daqueles que consideram o quadro normativo presente; e c) o de quem dirige sua busca para o futuro em inovações jurídicas.

Em primeiro lugar, convém destacar a inusitada proposta formulada por alguns autores no sentido de promover a "repristinação" de modelos históricos já superados, como a escravidão e a servidão[156]. No primeiro caso, ostentariam, pois, o *status* de objeto e não sujeito de direito. No segundo, seriam providos de alguns poucos direitos condicionados ao desempenho das atividades para as quais teriam sido desenvolvidos, à semelhança dos servos que possuíam algumas prerrogativas relativas ao uso da terra de seus senhores, como se recorda.

Há, ainda, os que pugnam pelo tratamento de tais agentes como o de meras ferramentas, desprovidos de qualquer traço de intencionalidade, de forma que toda e qualquer responsabilidade seria atribuída à pessoa (humana) sua proprietária ou usuária[157]. É a via adotada pelo ordenamento jurídico norte-americano e canadense, por ora[158].

De outro lado, podem ser mencionados os que sugerem a aplicação de soluções já presentes em diversos ordenamentos jurídicos, sustentando a utilização da analogia como ferramenta para empregar institutos hodiernamente positivados na legislação cível.

Nesse terceiro conjunto encontrar-se-iam os defensores das seguintes teorias: a) "representante" ("*the agent as a representative*") – segundo a qual os agentes atuariam em nome do "dono" do artefato (investido de uma espécie de mandato), o que implica certa dificuldade no tocante à imputação da vontade; b) "mero emissário" ("*the agent as a messenger*") - meros portadores de declaração de vontade alheia, os agentes não deveriam intervir na fixação do conteúdo da declaração (o que, entretanto, ocorreria na prática); e c)

---

[156] SCHWEIGHOFER, E – *Voruberlegungen zu kunstlichen Personen: autonome Roboter und intelligente Softwareagenten*. *Apud* WETTIG, Steffen and ZEHENDNER, Eberhard - *A legal analysis of human and electronic agents*, p. 127.

[157] É o que advoga JEAN-FRANÇOIS LEROUGE, consoante o qual: "*if a party creates a situation in which an electronic agent is to act on his behalf, then a party is bound by the actions of the agents*". Na mesma linha, WEITZENBOECK afirma que: "*the operations of an intelligent agent are attributed to the human who uses the agent*". Prossegue esse ultimo ressaltando que "*by initiating the electronic agent, the user is deemed to have accepted that contracts concluded by the agent will be binding on such user. The assent of the electronic agent will be inferred to be the assent of the (human) user of the agent*". *Apud* ANDRADE, Francisco & NOVAIS, Paulo & NEVES, José – *Op. cit.*, p. 5.

[158] ANDRADE, Francisco & NOVAIS, Paulo & NEVES, José – *Op. cit.*, p. 6.

"menor" ("*the agent as a minor*") – dotados de capacidade de contratação limitada, os agentes não agiriam por conta própria (isso importaria na subversão do instituto) [159].

Por fim, constata-se a existência daqueles que sustentam a tese segundo a qual seria imperioso desenvolver um novo modelo para regular esses novos sujeitos de direito. Advoga-se, assim, a criação de uma nova "personalidade jurídica", uma espécie de "*tertium genus*", entre as pessoas individuais (ou físicas, na terminologia do Direito Brasileiro) e coletivas (ou jurídicas). É o caso dos que pugnam pela regulação normativa de uma "*e-Person*" peculiar aos agentes artificiais, a exigir um registro eletrônico específico, a constituição de um fundo com recursos destinados a suportar as consequências econômicas dos atos por eles praticados[160], assim como a definição de uma responsabilização limitada ("Ltd. Agent")[161].

Cumpre notar que essa última corrente inspirou o Parlamento Europeu a propor o reconhecimento de um estatuto legal específico para entidades algorítmicas (e robôs)[162]. Após ter sido objeto de diversas críticas, a proposta foi rejeitada, ao menos por ora[163].

---

[159] WETTIG, Steffen and ZEHENDNER, Eberhard - *A legal analysis of human and electronic agents*, pp. 124-126.

[160] Uma das questões levantadas pelos críticos de tal proposta concerne à ausência de estímulo às "máquinas" quanto à preservação de sua existência enquanto "entes econômicos". É o que pondera FRANZISKA JANDL: "*One of the problems they see is that machines, unlike human beings, lack incentive to safeguard their economic existence and not to put their amount of liable capital at risk*". JANDL, Franziska - *E-Person – Treating robots like legal entities?*, p. 2.

[161] ZANKL, W. - *Juristische Aspekte Ku€nstlicher Intelligenz*. *Apud* WETTIG, Steffen and ZEHENDNER, Eberhard – *Op. cit.*, pp. 127-128.

[162] De acordo com o art.º 59.º a Proposta de Resolução formulada pelo Parlamento Europeu, contendo recomendações à Comissão sobre disposições de Direito Civil sobre Robótica (2015/2103): "59. Insta a Comissão, ao efetuar uma avaliação de impacto do respetivo futuro instrumento legislativo, a explorar, analisar e considerar as implicações de todas as soluções jurídicas possíveis, tais como: a) Criar um regime de seguros obrigatórios, se tal for pertinente e necessário para categorias específicas de robôs, em que, tal como acontece já com os carros, os produtores ou os proprietários de robôs seriam obrigados a subscrever um seguro para cobrir os danos potencialmente causados pelos seus robôs; b) Garantir que um fundo de compensação não serviria apenas para garantir uma compensação se um dano causado por um robô não se encontrasse abrangido por um seguro; c) Permitir que o fabricante, o programador, o proprietário ou o utilizador beneficiassem de responsabilidade limitada se contribuíssem para um fundo de compensação, bem como se subscrevessem conjuntamente um seguro para garantir a indemnização quando o dano é causado por um robô; d) Decidir quanto à criação de um fundo geral para todos os robôs autónomos inteligentes ou quanto à criação de um fundo individual para toda e qualquer categoria de robôs e quanto à contribuição que deve ser paga a título de taxa pontual no momento em que se coloca o robô no mercado ou quanto ao pagamento de contribuições periódicas durante o tempo de vida do robô; e) Garantir que a ligação entre um robô e o respetivo fundo seja patente pelo número de registo individual constante de um registo específico da União que permita que qualquer pessoa que interaja com o robô seja informada da natureza do fundo, dos limites da respetiva responsabilidade em caso de danos patrimoniais, dos nomes e dos cargos dos contribuidores e de todas as outras informações relevantes; f) Criar um estatuto jurídico específico para os robôs a longo prazo, de modo a que, pelo menos, os robôs autónomos mais sofisticados possam ser determinados como detentores do estatuto de pessoas eletrónicas responsáveis por sanar quaisquer danos que possam causar e, eventualmente, aplicar a personalidade eletrónica a casos em que os robôs tomam decisões autónomas ou em que interagem por qualquer outro modo com terceiros de forma independente; [...]". Atente-se, particularmente, para a alínea f do citado dispositivo.

[163] Informação disponível em: https://europa.eu/rapid/press-release_IP-18-3362_pt.htm

Entre as objeções levantadas, podem ser mencionadas: a) a atribuição de personalidade jurídica a tais entes artificiais resultaria no reconhecimento de direitos próprios de seres humanos, o que resultaria no esvaziamento da proteção conferida pelos diplomas internacionais aos direitos humanos; e b) semelhante estratégia não resolveria a questão da imputação de responsabilidade, pois os beneficiários da atuação dos entes eletrônicos seriam, ao fim, pessoas[164]. Houve, ainda, quem levantasse óbices formais, tais como a incompetência da Comissão ou Parlamento Europeus para definir os critérios a partir dos quais se poderia definir uma pessoa enquanto sujeito de direitos, matéria que seria prerrogativa da legislação nacional de cada Estado membro[165].

Seja como for, os contratos desenvolvidos e implementados por tais agentes desafiam as clássicas teorias contratuais, sobretudo em virtude do componente da imprevisibilidade própria da inteligência artificial. Entre outros aspectos, não se pode definir de antemão os termos da oferta feita pelo sistema, nem tampouco suas reações às eventuais tratativas desenvolvidas a partir daí[166].

Remanesce com preocupação central a tentativa de oferece alguma sorte de proteção quanto à capacidade de os indivíduos perseguirem objetivos razoáveis por meio de acordos confiáveis[167].

---

[164] A esse propósito, vale mencionar a Carta Aberta publicada em http://www.robotics-openletter.eu/

[165] No Canadá o debate foi travado, entre outros fóruns, no Grupo de Trabalho de Comércio Eletrônico nos seguintes termos, segundo o relato de KERR: "*I remember the very spirited discussion that @johndgregory, me and other members of the ULCC E-commerce Working Group s would have about nomenclature for the software/hardware systems that made such transactions possible. Ultimately we chose the term "electronic agent." A subsequent debate ensued about whether these electronic agents should be treated as mere instruments or as something more. In terms of our Working Group's ultimate recommendations for Ontario's Electronic Commerce Act, I was on the winning side of the first debate but lost on the second. In a later article (Spirits in the Material World), I elaborated on my position, arguing that a day would soon come when it would become pragmatic to think of electronic agents as exhibiting an intermediate ontology best dealt with by well established principles in the law of agency (an idea subsequently developed with much greater precision and care by Chopra and White)*". KERR, Ian - *The Arrival of Artificial Intelligence and "The Death of Contract"*, p. 3.

[166] É o que observa KERR: "*Even if we were to decide on policy grounds to attribute contractual liability to individuals who use or program these bots, those individuals are not "offerors" in any known sense of that word. How could they be? Those individuals have no clue if or when the system will make an offer, no idea what the terms of the offer will be, and no easy means of affecting the negotiations or trade once underway. AI-generated contracts of this sort problematize contract theory. I wonder whether Gilmore would view them as yet another nail in Contract's coffin?")*". KERR, Ian – *Op. cit.*, p. 6.

[167] BELLIA, Anthony J. – *Contracting with Electronic Agents*, p. 1.092.

CAPÍTULO II

2. Implicações concretas e potenciais quanto aos direitos humanos resultantes da implementação de sistemas de inteligência artificial

De acordo com o Comissariado de Direitos Humanos do Conselho da Europa, "o impacto da Inteligência Artificial (IA) nos direitos humanos é um dos fatores mais cruciais que definirão o período em que vivemos"[168].

É o que resulta do reconhecimento da presença ubíqua da tecnologia no cotidiano pessoal e profissional de pessoas individuais, assim como coletivas, tanto no âmbito privado, quanto na seara pública. Nesse último ponto, o aludido órgão europeu ressalta a utilização por parte de autoridades públicas para diversos fins, entre os quais a avaliação da personalidade e habilidades de seus cidadãos, a alocação de recursos financeiros. Trata-se de decisões suscetíveis de acarretar graves consequências para os direitos humanos[169].

A análise dos riscos emergentes dos sistemas de IA revela, pois, a existência de potencial de lesão a vários direitos humanos, a ser agravada pela "escalabilidade" e "ubiquidade" próprias de tais tecnologias. Não se cuida de uma mera preocupação especulativa a respeito de cenários futuros de longo prazo, em situações fartamente exploradas sobretudo na literatura e no cinema voltados para a ficção científica.

Na realidade, boa parte dos casos a seguir apresentados concerne às tecnologias já disponíveis, em grande medida, em diversos países.

2.1. Vida, integridade física e dignidade

Não se pode deixar de mencionar, de início, a existência de notáveis benefícios decorrentes do uso da IA em prol da defesa da vida, sobretudo para a melhoria das condições sanitárias, assim como na prevenção de surtos epidêmicos, no diagnóstico ou tratamento de

[168] CONSELHO DA EUROPA, Comissariado de Direitos Humanos - *Unboxing Artificial Intelligence: 10 steps to protect Human Rights*, p. 5.

[169] Nesse contexto, segundo o conjunto de recomendações formuladas pelo Comissariado, o objetivo central deveria ser o de assegurar "o equilíbrio certo entre desenvolvimento tecnológico e proteção dos direitos humanos". CONSELHO DA EUROPA, Comissariado de Direitos Humanos – *Op. cit.*, p. 5.

doenças[170], na mitigação das consequências das mudanças climáticas, entre outras possibilidades[171].

De outro lado, já não se pode ignorar a existência, mesmo no curto prazo, de ameaças relevantes à vida ou integridade física derivadas do uso de sistemas artificialmente inteligentes.

Um dos mais óbvios exemplos seria o das armas letais autônomas ("*autonomous weapon systems – AWS*"), a desafiarem os estudiosos do Direito Internacional, em virtude de questões como: i) riscos de danos colaterais a civis, com potencialidade de provocar a escalabilidade do conflito; ii) dificuldade de conciliação com as regras de Direito Humanitário, sobretudo no tocante à proteção a ser conferida aos não combatentes; e iii) o problema ético emergente da delegação para máquinas de decisões sobre vida e morte de pessoas[172].

Ainda nessa seara, tem-se também considerado o emprego da IA na assistência aos estrategos e outros atores no processo de planeamento e execução de operações bélicas, em um cenário de forte transferência das instâncias decisórias humanas a sistemas dotados de algum nível de inteligência artificial, com desdobramentos ainda imprevistos[173].

Há de se mencionar, outrossim, o desenvolvimento e implementação de tecnologias capazes de prover autonomia a veículos ou máquinas pesadas, as quais também representam risco importante de vulneração da incolumidade física das pessoas.

---

[170] Ilustrativo, a propósito, o feito alcançado por investigadores do Massachusetts Institute of Technology (MIT) que, no início de 2020, descobriram um novo antibiótico (denominado "halicina", em referência ao computador "Hal" do filme "2001, uma Odisseia no Espaço") para combater bactérias até aqui invulneráveis a outros medicamentos, graças à ajuda da IA. KISSINGER, Henry A.; SCHMIDT, Eric & HUTTENLOCHER, Daniel – A Era da Inteligência Artificial, pp. 15-17.

[171] ACCESS NOW – *Human Rights in the Age of Artificial Intelligence*, pp. 14-15.

[172] *INTERNATIONAL COMMITTEE OF THE RED CROSS (IRCR) – ICRC Position on Autonomous Weapon Systems*, p. 2.

[173] A esse respeito, pondera HENRY KISSINGER; "O efeito mais revolucionário e imprevisível poderá estar justamente no ponto em que IA e inteligência humana se encontram. Historicamente, os militares que planeiam uma batalha têm algum entendimento, ainda que lacunar, da doutrina, das táticas e da psicologia estratégica do adversário. Estes fundamentos e abordagens humanas comuns têm permitido o desenvolvimento de estratégias e táticas de confronto, bem como uma linguagem simbólica de ações militares demonstrativas, como interceptar um jato junto a uma fronteira ou navegar um navio numa rota marítima disputada. Todavia, quando uma força militar recorre à IA para planear ou selecionar alvos – ou até para assistência dinâmica durante uma patrulha ou um conflito -, estes conceitos e interações familiares poderão tornar-se indecifráveis, visto envolverem comunicação com, e interpretação de uma inteligência cujos métodos e táticas são uma incógnita". KISSINGER, Henry A.; SCHMIDT, Eric & HUTTENLOCHER, Daniel – A Era da Inteligência Artificial, p. 162.

No caso dos carros autónomos[174], as estatísticas não são consensuais, mas há estudos a indicarem uma elevada incidência de acidentes, embora sua gravidade costume ser menor do que as registadas naqueles conduzidos por motoristas humanos[175].

É no horizonte de longo prazo, contudo, que repousariam as ameaças mais sensíveis aos direitos em questão.

Embora haja uma imensa dificuldade em se delinear o cenário futuro da IA, diversos autores abordam as possibilidades "apocalípticas" relativas aos riscos globais de catástrofes resultantes do desenvolvimento da tecnologia[176].

Vale notar que a antropomorfização representa uma fonte de importante distorção cognitiva, a obstar, por vezes, a identificação dos reais perigos nesse âmbito[177]. A tendência de projetar nos sistemas de IA características (positivas e negativas) próprias dos seres humanos pode obnubilar o reconhecimento das potencialidades deletérias de tais agentes artificialmente inteligentes, sobretudo no tocante à vida e integridade física das pessoas.

Inspirados talvez nos estereótipos forjados pela ficção científica, alguns tendem a imaginar que uma inteligência artificial dotada de alta capacidade de processamento de informações deveria se comportar como um déspota provido de poderes ilimitados.

Entretanto, talvez a maior fonte de impactos potencialmente negativos resida no denominado "risco de desalinhamento" de valores ou objetivos[178], que seria especialmente

---

[174] De acordo com a SAE International ("*Society of Automotive Engineers*"), haveria, essencialmente, cinco níveis de automação: 1) "Driver Assistance" – provido de dispositivos de mera assistência ao motorista, como aceleração e câmbio (v.g. piloto automático adaptativo); 2) "*Partial Driving Automation*" – oferece controle de direção, aceleração e desaceleração, mas com a possibilidade de o motorista assumir o comando (v.g. piloto automático da Tesla); 3) "*Conditional Driving Automation*" (ou "sem olhos" – dispõe de mecanismos de detecção de dados ambientais (v.g. carros em movimento lento) que podem levar a decisões sobre todos os aspectos da direção (v.g. A8 de 2019 da Audi), mas obriga que o motorista se mantenha a disposição para realizar alguma intervenção; 4) "*High Driving Automation*" (ou "mind-off") – dispensa qualquer necessidade de atuação do motorista, mas apenas em determinada áreas específicas (v.g. Waymo do Google); 5) "*Full Driving Automation*" – prescindem de qualquer interação humana (v.g. Audi Aiconr5r5). Fontes: https://www.synopsys.com/automotive/autonomous-driving-levels.html e https://www.pocket-lint.com/pt-br/carros/noticias/143955-sao-explicados-os-niveis-de-direcao-autonoma [Consult. 23 mar. 2022]

[175] Segundo artigo publicado na *National Law Review*, haveria 9,1 acidentes por milhão de milhas dirigidas no caso de carros autônomos, em contraste com os 4,1 acidentes por milhão de milhas dirigidas em relação aos veículos conduzidos por pessoas. CLIFFORD LAW – *The Dangers of Driverless Cars – The National Law Review, January, 27, 2022, Volume XII, Number 27*. Para um estudo mais aprofundado acerca da matéria: PETROVIĆ, Đorđe & MIJAILOVIĆ, Radomir & PEŠIĆ, Dalibor Traffic – *Accidents with Autonomous Vehicles: Type of Collisions, Manoeuvres and Errors of Conventional Vehicles' Drivers*.

[176] Entre outros, podem ser citado os seguintes estudos: BOSTROM, Nick – *Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards*; e YUDKOWSKY, Eliezer – *Artificial Intelligence as a Positive and Negative Factor in Global Risk*.

[177] YUDKOWSKY, Eliezer – *Artificial Intelligence as a Positive and Negative Factor in Global Risk*, pp. 2-6.

[178] Na síntese de MATHIAS RISSE: "*The pertinent challenge is the problem of value alignment, a challenge that arises way before it will ever matter what the morality of pure intelligence is. No matter how precisely AI systems are generated we must try to make sure their values are aligned with ours to render as unlikely as possible any complications from the fact that a superintelligence might have value commitments very different from ours*". RISSE, Mathias – *Human Rights and Artificial Intelligence*, pág. 8.

grave quando e se alcançada a "Super Inteligência Artificial" ("*Artificial Super Intelligence - ASI*")[179] ou mesmo a "Inteligência Artificial Geral" ("*Artificial General Intelligence – AGI*"), o que teria o potencial de deslocar o eixo de poder das instituições "humanas" existentes para agentes artificialmente inteligentes.

Trata-se de risco derivado das imensas dificuldades envolvidas no processo de assimilação por parte dos modelos de IA dos valores, necessidades e interesses humanos[180], ao definir os meios necessários ao atingimento das metas fixadas. O desalinhamento referido concerniria, pois, à possibilidade de ocorrer o desencontro ou falha de coincidência entres as metas perseguidas pelos agentes artificiais e aquelas buscadas pelos seres humanos.

A gravidade de tal risco tem sido qualificada como de natureza existencial[181]. A criação de uma "superinteligência" poderia levar à aniquilação da espécie humana, em face de equívocos derivados do aludido desalinhamento[182]. Poder-se-ia, ilustrativamente, elevar o status de uma "submeta" ao de "supermeta" – v.g. para resolver um problema matemático o agente faz transforma toda a matéria no sistema solar em um gigantesco aparelho de calcular, eliminando toda a vida no planeta[183].

Para melhor compreender a questão, também se poderia fazer um exercício hipotético a partir das metas fixadas pelos seres humanos em detrimento de outras espécies vivas. Ao se construir uma represa, exemplificativamente, não se costuma dispender muito tempo a considerar os seus impactos na destruição de colônias de insetos que porventura habitem na

---

[179] Alguns teóricos avaliam como inviável a criação de mecanismos de contenção de riscos de uma ainda hipotética superinteligência articial: "*Superintelligence is a hypothetical agent that possesses intelligence far surpassing that of the brightest and most gifted human minds. […] We argue that total containment is, in principle, impossible, due to fundamental limits inherent to computing itself. Assuming that a superintelligence will contain a program that includes all the programs that can be executed by a universal Turing machine on input potentially as complex as the state of the world, strict containment requires simulations of such a program, something theoretically (and practically) impossible*" – ALFONSECA, Manuel & CEBRIAN, Manuel & ANTA, Antonio Fernández & COVIELLO, Lorenzo & ABELIUK, Andrés & RAHWAN, Iyada - *Superintelligence Cannot be Contained: Lessons from Computability Theory*, p. 1.

[180] A proposito do tema, entre outros: CHRISTIAN, Brian – *The Alignment Problem – How Can Artificial Intelligence Learn Human Values*.

[181]De acordo com um dos teóricos a se debruçarem com mais profundidade sobre o assunto, NICK BOSTROM, risco existencial poderia ser definido como "*One where an adverse outcome would either annihilate Earth-originating intelligent life or permanently and drastically curtail its potential*". Dentre as hipóteses enunciadas pelo autor encontram-se riscos resultantes do mau uso da nanotecnologia, holocausto nuclear, agentes biológicos modificados. BOSTROM, Nick - *Existential Risks - Analyzing Human Extinction Scenarios and Related Hazards*, p. 2.

[182] BOSTROM, Nick – *Op. cit.*, p. 5.

[183] Embora pueril, o exemplo formulado BOSTROM ilustra a dinâmica do problema do desalinhamento antes referido.

zona em vias de ser inundada. O mesmo, em tese, poderia vir a ocorrer quanto à execução de metas por parte de uma Superinteligência artificial[184].

Note-se que aqui não se coloca o problema da "autoconsciência" como premissa para a elevação dos riscos supramencionados, diversamente do que poderia sugerir uma visão antropomorfizada desse cenário especulativo futuro. Isso porque o desalinhamento poderia resultar de defeito na fixação ou assimilação das metas, sem qualquer correlação com a existência de algum nível de autopercepção por parte dos agentes superinteligentes.

Seja como for, ainda que se possa considerar baixa a probabilidade de se concretizarem essas ameaças, mostra-se positivo o "custo-benefício" de se investirem esforços no trabalho de contenção, dada o nível de risco envolvido[185].

2.2. Pluralismo

A definição de uma base comum ("*common ground*") de valores a serem tutelados na regulação da IA representa um dos maiores desafios aos que pretender se debruçar sobre o problema.

A principal causa refere-se à própria diversidade de perfis dos atores que figuram como protagonistas na criação de tais sistemas. Trata-se de entes privados e públicos com agendas específicas de interesses, assim como contextos socioeconômicos e culturais profundamente heterogêneos[186].

Há quem questione não só a viabilidade de uma abordagem que busque encontrar um consenso para formação de um núcleo duro de princípio éticos a orientarem a construção dos marcos regulatórios. Para esses, seria o caso de abraçar como premissa não apenas tolerável, mas desejável, a diversidade própria às comunidades humanas[187].

---

[184] O exemplo é de MAX TEGMARK e encontra-se assim apresentado: "*people don't think twice about flooding anthills to build hydroelectric dams, so let's not place humanity in the position of those ants*". TEGMARK, Max - *Life 3.0 – Being Human in the Age of Artificial Intelligence*, p. 334.

[185] É o que recorda BOSTROM, ao afirmar que: "*Although there is still a fairly broad range of differing estimates that responsible thinkers could make, it is nonetheless arguable that because the negative utility of an existential disaster is so enormous, the objective of reducing existential risks should be a dominant consideration when acting out of concern for humankind as a whole. It may be useful to adopt the following rule of thumb for moral action; we can call it Maxipok: Maximize the probability of an okay outcome, where an "okay outcome" is any outcome that avoids existential disaster*". BOSTROM, Nick – *Op. cit.*, p. 20.

[186] RUDSCHIES, Catharina & SCHNEIDER, Ingrid & SIMON, Judith - *Value Pluralism in the AI Ethics Debate – Different Actors, Different Priorities*.

[187] Eis como defendem a tese RUDSCHIES, SCHNEIDER e SIMON: "*As the diverging preferences of actor types as well as the varying sets of synthesised principles of the other reviews reveal, it proves rather difficult to find one universal set of principles in the context of AI. This is especially true in a pluralist society, where individuals and subgroups have different backgrounds and, thus, value perspectives on the issue at hand. Hence, one must refute a monist view of an ethics of AI, which assumes that there is a single value or a small set of values*

O pluralismo colocar-se-ia, nesse contexto, como um "valor em si" a ser protegido.

Entretanto, as características inatas ao processo de desenvolvimento dos sistemas de IA parecem indicar a imensa dificuldade em se implementar uma abordagem regulatória pluralista ou que incorpore o pluralismo como valor intrínseco.

Dentre os aspetos mais destacados, pode-se mencionar a acentuada massificação resultante do tratamento e manipulação das informações por parte de agentes artificialmente inteligentes.

Basta uma rápida análise dos padrões de comportamento emergentes da exposição iterativa às plataformas digitais mais populares (v.g. Instagram, Facebook e Twitter) para constatar o efeito deletério à pluralidade que decorre do forte emprego de tecnologias assistidas por IA, ao induzir tendências de consumo[188], inclinações políticas[189] ou mesmo orientações morais[190].

Emerge desse cenário um grave risco de homogeneização do comportamento humano, de modo a suprimir traços individualizantes e, assim, comprometer as diversidades ideológicas, políticas e culturais que marcam o desejável pluralismo como direito fundamental.

---

that overrides all others or provides guidance on how to compare one value with another. Rather, we need to acknowledge that the AI ethics landscape is characterised by a plurality of sets of values. According to the theory of moral or value pluralism, there is, due to diverging understandings of a good life, a multitude of genuine human goods and values". RUDSCHIES, Catharina & SCHNEIDER, Ingrid & SIMON, Judith – *Op. cit.*, p.9.

[188] Sobre a eficiência do uso da IA como ferramenta para ampliar o potencial de "*marketing*" das redes sociais, entre outros: SANTY, R D et al. - *Artificial Intelligence as Human Behavior Detection for Auto Personalization Function in Social Media Marketing.*

[189] Graças sobretudo à potencialização resultante do uso massivo de recursos de IA, o peso exercido pelas empresas que gerem as plataformas digitais nos processos democráticos levou o Conselho da Europa a adotar Recomendação aos Estados membro acerca de sua responsabilidade, nesse contexto. Entre outros pontos, eis o que se enfatizou em discurso dirigido aos Ministros responsáveis pela "Mídia e Sociedade da Informação" no aludido conselho: "*The use of AI systems in filtering online content has led to restrictions of user-generated content even at the point of upload and enabled the spread of disinformation. It is usually stressed that freedom of expression should be equally protected online and offline. However, the role played by internet intermediaries online has from the beginning blurred the lines. Companies operating social media platforms have such control over access to information that they can shape public debate with little accountability, if any. Therefore, any decisions that these companies make can have huge implications for our democracies. This has been recognised early on by the Committee of Ministers of the Council of Europe, which adopted a Recommendation to member States on the roles and responsibilities of internet intermediaries in 2018*" (Comissariado para Direitos Humanos - *Artificial Intelligence – Intelligent Politics. Challenges and opportunities for media and democracy*).

[190] A respeito dos riscos de manipulação do comportamento humano por sistemas de IA, entre outros: PETROPOULOS, Georgios - *The dark side of artificial intelligence: manipulation of human behaviour*.

2.3. Inclusão

Tecnologia e inclusão mantém uma realização de coexistência simbiótica, de um lado, mas também de algum antagonismo, de outro.

Os desdobramentos das revoluções industriais figuram como um ilustrativo exemplo nesse sentido, ao evidenciarem, de uma parte, as diversas consequências positivas para o desenvolvimento económico e social e os grandes desafios para manter o acesso igualitário às oportunidades geradas, de outra.

No caso da IA, o mesmo fenómeno se verifica.

Há, exemplificativamente, um enorme potencial inclusivo no seu uso para promover o aprimoramento da eficiência econômica de empresas de pequeno porte[191].

Entretanto, subsiste uma quase insuperável fonte de desequilíbrio advinda da própria heterogeneidade constatada nos estágios de desenvolvimento em que se encontram regiões, países, empresas, instituições académicas e indivíduos quanto ao acesso às tecnologias que contam com IA "embarcada".

Apenas para se aquilatar as óbvias disparidades existentes entre as nações, vale ter em conta a quarta edição do ranking elaborado pela Universidade de Oxford em 2021, que leva em conta 42 indicadores e 10 dimensões, para medir o nível de preparo dos governos 160 países para implementação de soluções de IA destinadas aos seus cidadãos[192]. Abaixo a relação dos 10 primeiros colocados, de um lado, e os 10 últimos, de outro:

[191] A esse propósito, TIMOTHY KTOIN, especialista da OCDE, oferece um interessante exemplo do uso de ferramentas de IA para potencializar as oportunidades de negócios em pequenas empresas africanas: "*The use of data analytics and artificial intelligence (AI) technologies can help overcome some of these barriers and re-balance opportunity. Superfluid Labs focus on empowering hundreds of small businesses and millions of individuals across Africa with technology. The goal is to unlock new economic opportunities where there were none and expand existing avenues where they are limited. Business efficiency through better data insights yields many dividends for small businesses, such as the millions of informal merchants who employ most of Africa's increasingly youthful workforce. For example, informal merchants in Nigeria and Kenya can now better predict consumer demand for their goods and easily place orders for new stock via basic mobile devices. Additionally, by digitising their sales transactions for the first time, many of these businesses can access loans from local lenders through the alternative credit scores Superfluid Labs generate using artificial intelligence. The result is enhanced business viability and expanded employment capacity, which in turn attract more risk capital for new business formation*". KOTIN, Timothy - *Artificial intelligence and data analytics can unlock new economic opportunities*.
[192] OXFORD INSIGHTS - *Government AI Readiness Index 2021*, págs. 61 e 68.

Tabela 2.

Classificação de nações quanto à implementação – 10 primeiros colocados[193]

Global Position	Country	Overall Score	Government	Technology Sector	Data and Infrastructure
1	United States of America	88.16	88.46	83.31	92.71
2	Singapore	82.46	94.88	66.69	85.80
3	United Kingdom	81.25	85.69	67.26	90.81
4	Finland	79.23	88.45	63.85	85.40
5	Netherlands	78.51	80.42	66.17	88.92
6	Sweden	78.16	80.76	67.37	86.36
7	Canada	77.73	84.36	63.75	85.08
8	Germany	77.26	78.04	67.68	86.07
9	Denmark	76.96	83.50	63.24	84.14
10	Republic of Korea	76.55	85.27	58.49	85.89

Tabela 3.

Classificação de nações quanto à implementação – 10 últimos colocados[194]

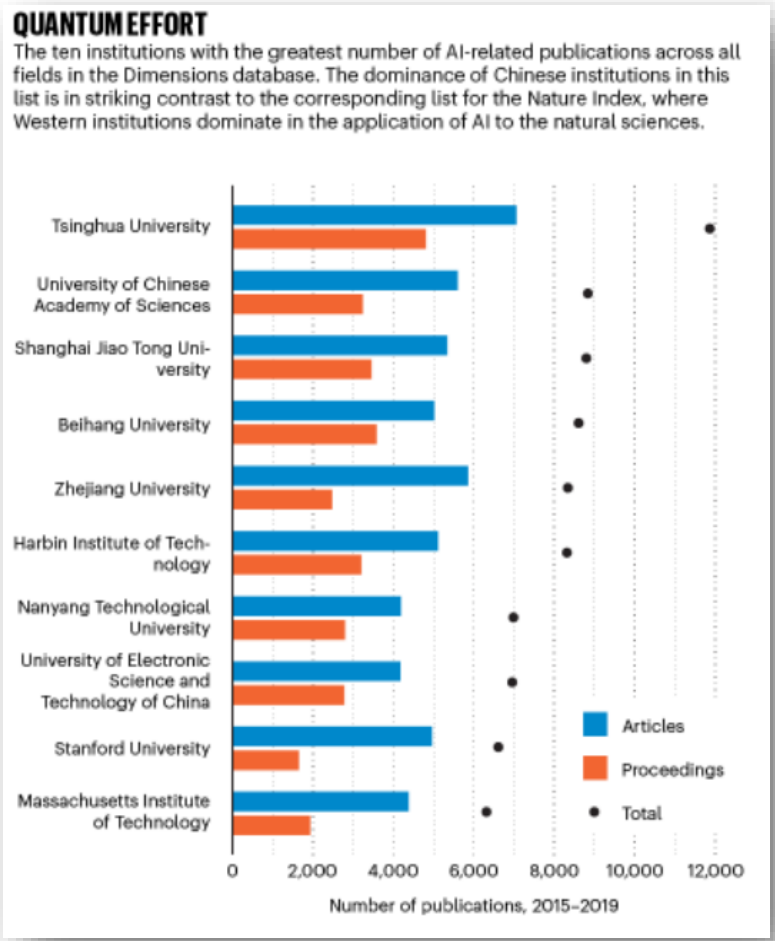
Global Position	Country	Overall Score	Government	Technology Sector	Data and Infrastructure
151	Sudan	25.91	24.83	21.62	31.29
152	Haiti	25.14	24.24	14.67	36.49
153	Malawi	24.85	28.57	18.43	27.55
154	Afghanistan	24.38	25.67	9.23	38.23
155	Chad	24.00	27.58	14.41	30.02
156	Burundi	23.72	28.40	19.06	23.70
157	Democratic Republic of the Congo	23.32	27.46	16.72	25.79
158	Angola	22.87	27.30	13.86	27.44
159	Central African Republic	20.73	15.01	19.21	27.98
160	Yemen	17.93	15.85	17.91	20.03

[193] OXFORD INSIGHTS – *Op. cit.*, pág. 61.
[194] OXFORD INSIGHTS – *Op. cit.*, pág. 68.

Outro dado relevante para aferir o desnível já no campo acadêmico concerne à quantidade de publicações por universidade em relação à IA, entre os anos de 2015 e 2019. Em um estudo divulgado pela *Nature*, nota-se a grande concentração nas instituições chinesas e norte-americanas, com notável predomínio das primeiras:

Figura 3.

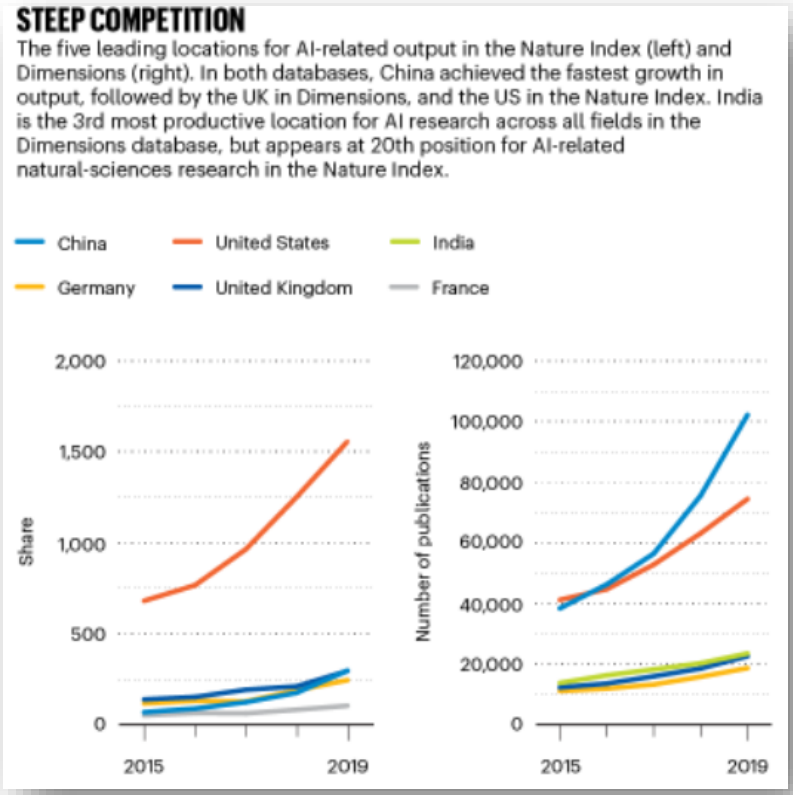
Índice de publicações acadêmicas sobre IA, por instituição de ensino[195]



[195] NATURE - *The race to the top among the world's leaders in artificial intelligence*. Para uma análise dessa disputa, sob a perspectiva estadunidense: CENTER FOR SECURITY AND EMERGING TECHNOLOGY (CSET) – *The Question of Comparative Advantage in Artificial Intelligence*.

Figura 4.

Índice de publicações acadêmicas sobre IA, por local[196]



Ora, dessa nada surpreendente desigualdade internacional emerge a constatação de que a inclusão representa um desafio especial nos países estruturalmente despreparados para incorporar tais tecnologias.

Aliás, as propriedades exponencialmente disruptivas de boa parte dos sistemas de IA podem vir a agravar essa abissal desproporção. Com efeito, os países (ou entes privados) que alcançarem níveis mais elevados de desenvoltura no uso de semelhantes ferramentas contarão com diferenciais competitivos, cujo potencial de eliminar a capacidade de concorrência dos demais pode ter efeitos devastadores quanto à inclusão nas suas mais diferentes perspectivas (social, econômica, cultural, etc.).

A mesma disparidade pode ser constatada, naturalmente, no âmbito interno dos países, nos quais se verifica alguma desigualdade econômica, social ou cultural, a repercutir no modo como a IA poderá impactar de modo heterogêneo a vida das pessoas sujeitas a condições de vida profundamente diversificadas.

---

[196] NATURE - *The race to the top among the world's leaders in artificial intelligence*. Para uma análise dessa disputa, sob a perspectiva estadunidense: CENTER FOR SECURITY AND EMERGING TECHNOLOGY (CSET) – *The Question of Comparative Advantage in Artificial Intelligence*.

2.4. Isonomia

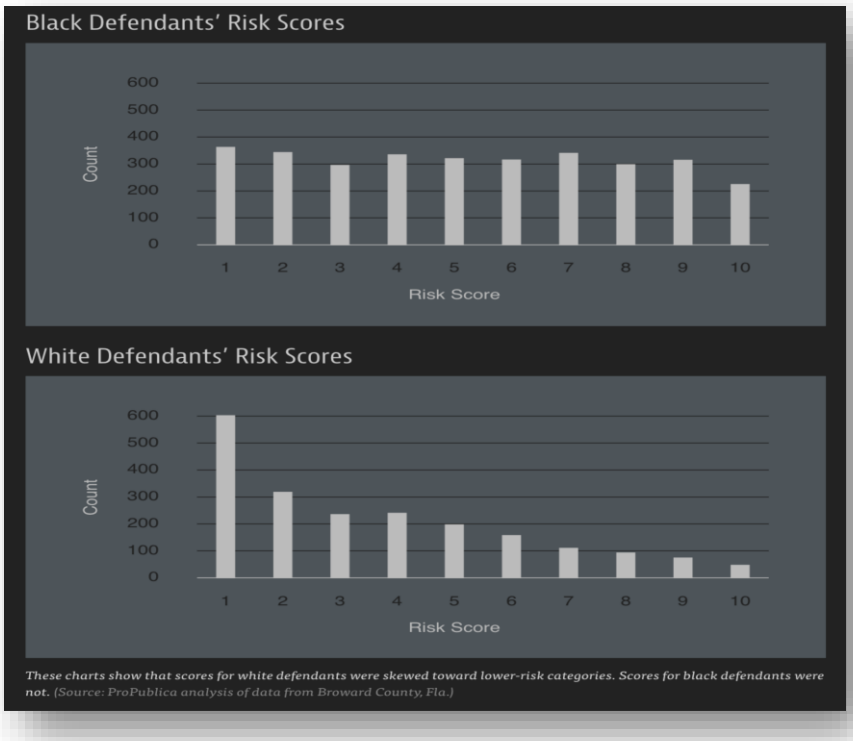
Uma das principais questões relativas à isonomia, na fase atual da IA, consiste no fato de as análises preditivas realizadas pelos sistemas hodiernos estarem centradas em avaliações atuariais fundadas em elementos estatísticos.

Conclusões fundadas na inferência de inclinações com lastro no comportamento pretérito de um indivíduo ou, o que é pior, do grupo de que faz parte podem produzir injustiças profundas.

Em um estudo instigante conduzido e divulgado pela *Pro Publica*[197] foram identificadas distorções que revelariam aparente enviezamento ("*machine bias*") racial de algoritmos utilizados para definição de indicadores de risco concernentes à reincidência criminal, quase como uma espécie de "*lombrosianismo algorítmico*". Eis a compilação estatística de acordo com o recorte étnico:

Figura 5.

Estatísticas da Pro Publica sobre o caso enviezamento racial no caso "Compass" [198]



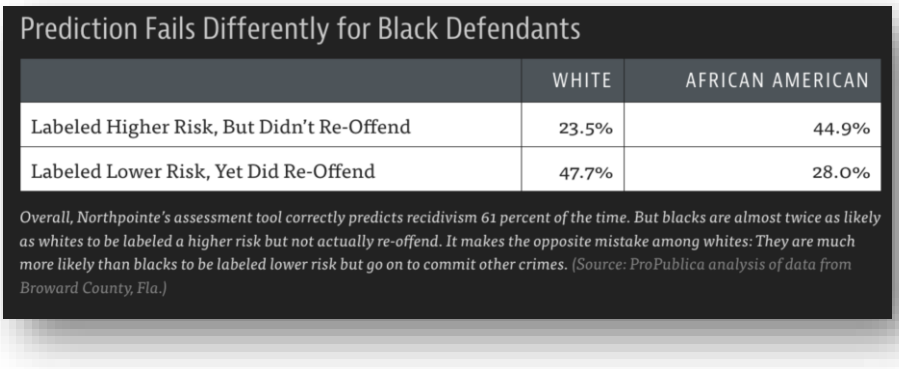
[197] De acordo com a autodescrição contida no seu sítio eletrônico, a Pro Publica seria "*um corpo jornalístico independente e sem fins lucrativos que produz jornalismo investigativo com força moral*". No original "*ProPublica is an independent, nonprofit newsroom that produces investigative journalism with moral force*". Informações disponíveis em https://www.propublica.org/about/ [Consult. 24 dez. 2018].

[198] Informações disponíveis em https://www.propublica.org/about/ [Consult. 24 dez. 2018].

Quanto às falhas de previsão, ao que parece, também haveria algum desiquilíbrio. Os dados seriam bastante elucidativos:

Tabela 4.

Estatísticas da Pro Publica sobre o falhas preditivas no caso "Compass" [199]



Prediction Fails Differently for Black Defendants

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from Broward County, Fla.)

Embora a investigação sobre as causas de tais supostas disfunções mereça aprofundamento, o estudo revela uma grave preocupação quanto aos parâmetros adotados pelos sistemas para apresentar suas conclusões.

Distorções semelhantes também podem ser identificadas em programas de recrutamento nos quais se emprega alguma tecnologia artificialmente inteligente, pensados, precisamente, para diminuir a subjetividade e o arbítrio no processo seletivo. No entanto, tem sido detectado algum nível de contaminação dos algoritmos, de forma a ensejar preferências e rejeições, com base em critérios não razoáveis[200].

Outro exemplo relevante em que há potencial lesivo à isonomia no uso de IA concerne aos sistemas para gestão de pontuações de crédito ("credit scores"), em que se pode restringir ou obstar o acesso a fontes de financiamento, a partir de análises enviazadas[201].

Não obstante, a antes referida opacidade das ferramentas de suporte à decisão impede a sua auditagem mais rigorosa a busca das reais origens dos seus alegados defeitos de funcionamento.

---

[199] Informações disponíveis em https://www.propublica.org/about/ [Consult. 24 dez. 2018].

[200] Um dos casos reportados foi o verificado no sistema utilizado pela Amazon, que estaria a privilegiar candidatos do sexo masculino nas seleções realizadas. HSU, Jeremy – *AI Recruiting Tools Aim to Reduce Bias in the Hiring Process*.

[201] Interessante exemplo citado por CATHY O´NEIL consiste no uso de "machine learning" para identificar padrões associados a determinados riscos de inadimplência no preenchimento de formulários para obtenção de crédito – v.g. o tempo médio de leitura dos textos e eventuais incorreções gramaticais ou ortográficas, a sinalizarem tratar-se de pessoas com baixa instrução ou estrangeiros. O´NEIL, Cathy - *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, p. 17.

Por conseguinte, mister definir mecanismos apriorísticos ("design isonômico" em linha com a ideia do "*human rights by design*") que permitam apurar e eliminar as falhas de onde resultem vulnerações anti-isonômicas.

Em último lugar, costuma ser lembrada como risco epistêmico das ferramentas de IA a possibilidade de contaminação arquitetural dos algoritmos, a ensejar uma espécie de enviesamento ("*algorithm bias*")[202].

Nesse contexto, o uso de "*data sets*" viciados poderia resultar, naturalmente, em graves falhas no processo decisório[203]. Importa salientar que os erros poderiam resultar não apenas da "transferência" de "desvios subjetivos" dos programadores para os algoritmos, mas também da ampliação das distorções por meio da reintrodução de dados contaminados no processo de aprendizado de máquina[204].

Podem ser citadas, entre as fontes principais de enviesamento cognitivo que poderiam macular sistemas artificialmente inteligentes, as que se seguem: a) "*bandwagon effect*" (efeito manada) - resultante de uma tendência a presumir ou suspender o juízo crítico quanto a premissas fixadas pelo pensamento coletivo; b) "*confirmation bias*" (*myside bias* - viés da confirmação) - uso tendencioso de informações para referendar inclinações apriorísticas, confirmando o que o indivíduo já supunha saber "desde o início"; e c) "*selective perception*" (percepção seletiva) - recorte seletivo de dados que se alinhem a determinadas tendências do sujeito cognoscente, repelindo ou abstraindo os demais elementos[205].

---

[202] BAROCAS, Solon and SELBST, Andrew D. - *Big Data's Disparate Impact*. Os autores ressaltam que os mesmos preconceitos presentes em processos decisórios humanos poderiam macular aqueles conduzidos por algoritmos. Em suas palavras: "*Advocates of algorithmic techniques like data mining argue that these techniques eliminate human biases from the decision-making process. But an algorithm is only as good as the data it works with. Data is frequently imperfect in ways that allow these algorithms to inherit the prejudices of prior decision makers. In other cases, data may simply reflect the widespread biases that persist in society at large*". BAROCAS, Solon and SELBST, Andrew D. - *Op. cit.*, p. 671.

[203] A esse propósito, vale recordar a diferenciação proposta por KAHNEMAN entre "bias" e "noise", fatores que podem comprometer, alternativa ou cumulativamente, processos decisórios. "*When people consider errors in judgment and decision making, they most likely think of social biases like the stereotyping of minorities or of cognitive biases such as overconfidence and unfounded optimism. The useless variability that we call noise is a different type of error. To appreciate the distinction, think of your bathroom scale. We would say that the scale is biased if its readings are generally either too high or too low. If your weight appears to depend on where you happen to place your feet, the scale is noisy. A scale that consistently underestimates true weight by exactly four pounds is seriously biased but free of noise. A scale that gives two different readings when you step on it twice is noisy. Many errors of measurement arise from a combination of bias and noise. Most inexpensive bathroom scales are somewhat biased and quite noisy"*. KAHNEMAN, Daniel & ROSENFIELD, Linnea Gandhi & BLASER, Tom - *Op. cit.*, p. 7.

[204] "*The issue is that bias is not only **transferred** to algorithms, but it can also be **amplified.** This happens when biased ML algorithms create new data that is re-injected into the model as part of its continuous training. It becomes worse when the biased algorithm is used to make millions of predictions per minute as part of an automatic decision-making process, bringing bias back into the real world, at scale*". PENEL, Olivier - *Algorithms, the Illusion of Neutrality: The Road to Trusted AI*, p. 3.

[205] BAROCAS, Solon and SELBST, Andrew D. - *Op. cit.*, p. 672.

Tais falhas poderiam ocorrer em diversos momentos do aprendizado profundo das máquinas, desde a descrição ou enquadramento do problema ("*framing the problem*"), passando pela coleta dos dados, até a sua preparação e manuseio[206]. A dificuldade em corrigi-las derivaria, entre outros fatores, da ausência de mecanismos seguros para assegurar a rastreabilidade dos desvios, da ausência de contextualização social ("*lack of social context*") a obstar a adequação dos modelos às situações nas quais serão aplicadas[207], assim como da definição de categorias abstratas de notável complexidade (v.g. conceitos de justiça, igualdade, liberdade, entre outros)[208].

Sobreleva notar que há diversas tentativas em curso para aprimorar os modelos matemáticos e calibrar os sistemas de aprendizado de máquinas para depurar aludidas distorções[209].

2.5. Privacidade

Sistemas de inteligência artificial costumam manusear quantidades massivas de dados e, em geral, sua eficiência depende do volume de informações presentes em sua base[210]. Por conseguinte, a sua extração tem uma importância estratégica inegável para a realização de boa parte de suas operações.

Até relativamente pouco tempo não havia grande sensibilidade quanto ao valor jurídico dos dados pessoais, razão pela qual não havia limites legais explícitos relativamente à possibilidade não apenas de coletar mas mesmo de comercializar tais informações[211].

Vale observar que, por vezes, a coleta de dados vista sob a perspectiva estrita do conteúdo específico e individual de informações "neutras" ou "inofensivas" obtidas pode conduzir a uma falsa sensação de inocuidade jurídica, por ausência de lesões a direitos de

---

[206] HAO, Karen - *This is how AI bias really happens—and why it's so hard to fix*, p. 1.

[207] Eis como a descrevem BAROCAS e SEBST, ao identificarem o que denominam "*portability trap*": "*Within computer science, it is considered good practice to design a system that can be used for different tasks in different contexts. But what that does is ignore a lot of social context. You can't have a system designed in Utah and then applied in Kentucky directly because different communities have different versions of fairness. Or you can't have a system that you apply for 'fair' criminal justice results then applied to employment. How we think about fairness in those contexts is just totally different*". *Apud* HAO, Karen - *Op. cit.*, p. 3.

[208] HAO, Karen - *Op. cit.*, p. 4.

[209] COWGILL, Bo - *Bias and Productivity in Humans and Algorithms: Theory and Evidence from Re sumé´s Screening*.

[210] Os dados costumam ser empregados para criar mecanismos de "feedback", de modo a permitir a calibragem e refinamento das ferramentas. ACCESS NOW – *Human Rights in the Age of Artificial Intelligence*, p. 20.

[211] No plano europeu, passou a ser aplicável o Regulamento Geral sobre a Proteção de Dados (RGPD) a partir de 25 de maio de 2018, um conjunto normativo destinado a uniformizar as regras sobre proteção de dados para todas as empresas ativas na UE, independentemente da sua localização. Já no Brasil, foi a Lei 13.709, de 14 de agosto de 2018, que se incumbiu de regular o tema.

personalidade, quando se pensa, ilustrativamente, em preferências gastronômicas, tempos de deslocamento ou períodos de inatividade.

Entretanto, tal como propõe a Teoria do Mosaico, a avaliação deve se dar a partir do conjunto dos dados capturados e da interrelação entre eles, v.g. para identificação de padrões de consumo, hábitos, vícios ou relações interpessoais, com o cruzamento de referências concernentes a localização, expressões de busca na Internet ou histórico de visitas a determinados sítios[212].

Ilustrativo, nessa esteira, o uso de aprendizado de máquina ("machine learning") para determinar genêro, estado civil e até idade, a partir da mera análise de dados relativos à geolocalização, os quais também permitiriam realizar uma avaliação preditiva da provável localização futura dos sujeitos investigados, em determinado horizonte temporal[213].

As ameaçam ganham exponencialidade ao se considerar a pletora de instrumentos de captura de dados disponíveis, para além dos "smartphones", tais como os eletrodomésticos inteligentes ("internet of things") conectados às redes 5G, veículos quase autônomos e as câmeras de vigilância cada vez mais potentes e sofisticadas, que integram o cenário hodierno das "*smart cities*"[214].

---

[212] "*Mosaic theorists advocate a cumulative approach to the evaluation of data collection. Under the theory, searches are 'analyzed as a collective sequence of steps rather than as individual steps'. The approach is based on the observation that comprehensive aggregation of even seemingly innocuous data reveals greater insight than consideration of each piece of information in isolation. Over time, discrete units of surveillance data can be processed to create a mosaic of habits, relationships, and much more. Consequently, a Fourth Amendment analysis that focuses only on the government's collection of discrete units of data fails to appreciate the true harm of long-term surveillance—the composite*". BELLOVIN, Steven M. & HUTCHINS, Renée M. & JEBARA, Tony & ZIMMECK, Sebastian – *When Enough is Enough: Location Tracking, Mosaic Theory, and Machine Learning*, p. 556.

[213] "*In a recent competition, 'The Nokia Mobile Data Challenge', researchers evaluated machine learning's applicability to GPS and cell phone tower data. From a user's location history alone, the researchers were able to estimate the user's gender, marital status, occupation and age.8 Algorithms developed for the competition were also able to predict a user's likely future location by observing past location history. The prediction of a user's future location could be even further improved by using the location data of friends and social contacts. the user's gender, marital status, occupation and age. Algorithms developed for the competition were also able to predict a user's likely future location by observing past location history. The prediction of a user's future location could be even further improved by using the location data of friends and social contacts*". BELLOVIN, Steven M. & HUTCHINS, Renée M. & JEBARA, Tony & ZIMMECK, Sebastian – *When Enough is Enough: Location Tracking, Mosaic Theory, and Machine Learning*, pp. 557-558.

[214] ACCESS NOW – *Human Rights in the Age of Artificial Intelligence*, p. 21.

O principal dilema aqui consiste em compaginar as vantagens econômicas e sociais resultantes da obtenção e manuseio de aludidos dados[215], de um lado, com a preservação de direitos individuais, particularmente a privacidade[216].

O conceito de autodeterminação informacional (ou informativa)[217] parece figurar no centro de gravidade das soluções aventadas para a resolução do impasse.

Derivam dele alguns dos princípios positivados pela legislação europeia de proteção de dados, entre os quais podem ser destacados:

> a) Objeto de um tratamento lícito, leal e transparente em relação ao titular dos dados («licitude, lealdade e transparência»); b) Recolhidos para finalidades determinadas, explícitas e legítimas e não podendo ser tratados posteriormente de uma forma incompatível com essas finalidades; o tratamento posterior para fins de arquivo de interesse público, ou para fins de investigação cientifica ou histórica ou para fins estatísticos, não é considerado incompatível com as finalidades iniciais, em conformidade com o artigo 89.o, n.o 1 («limitação das finalidades»); c) Adequados, pertinentes e limitados ao que é necessário relativamente às finalidades para as quais são tratados («minimização dos dados»); [...][218].

Outro desdobramento essencial consiste na exigência de expressa manifestação de consentimento como condição de validade para a condução de operações de retenção e manipulação de dados pessoais[219].

---

[215] Não se pode deixar de considerar que o avanço das novas tecnologias tem acarretado formas de "invasão de privacidade" nunca antevistas e até aqui percebidas como ficção científica, como é o caso da ressonância magnética funcional ("*functional magnetic resonance imaging*" - fMRI). Trata-se de nova abordagem que permite a leitura de ondas cerebrais para detecção de mentira (como acurácia de até 90%) ou mesmo a extração de informações mais específicas como pensamentos ou memórias a respeito da eventual subtração de um objeto (com precisão de cerca de 70%). LEWIS, Tanya & WRITER, Staff - *Brain Says Guilty! Neural Imaging May Nab Criminals*.

[216] É o que também ressaltam KAPLAN e HAENLEIN: "*A central issue in this context is the topic of privacy and data protection. In the near future, we will face a wide range of dilemmas that require balancing social advances in the name of AI and fundamental privacy rights*". Seguem ainda afirmando que os Estados se veriam na contingência de restringir a privacidade e progresso econômico, ressaltando que há certa heterogeneidade nas decisões tomadas pelos diversos países que estão enfrentando a questão: "*This puts States in the complex position to decide how much privacy can and should be sacrificed on the altar of economic growth. Different countries are already making different decisions in this respect and those decisions will likely shape AI trends in years and even decades to come*". KAPLAN, Andreas & HAENLEIN, Michael - *Op. cit.*, p. 5.

[217] O conceito figura como um dos fundamentos em que a Lei 13.709/18 fez repousar o regime brasileiro de proteção de dados, consoante se extrai do seu art.º 2.º, II. BRASIL - Lei 13.709/18, de 14/08/18.

[218] Em sentido semelhante podem ser lembrados os princípios agasalhados no ordenamento jurídico brasileiro: "I - finalidade: realização do tratamento para propósitos legítimos, específicos, explícitos e informados ao titular, sem possibilidade de tratamento posterior de forma incompatível com essas finalidades; II - adequação: compatibilidade do tratamento com as finalidades informadas ao titular, de acordo com o contexto do tratamento; III - necessidade: limitação do tratamento ao mínimo necessário para a realização de suas finalidades, com abrangência dos dados pertinentes, proporcionais e não excessivos em relação às finalidades do tratamento de dados; IV - livre acesso: garantia, aos titulares, de consulta facilitada e gratuita sobre a forma e a duração do tratamento, bem como sobre a integralidade de seus dados pessoais; [...] VI - transparência: garantia, aos titulares, de informações claras, precisas e facilmente acessíveis sobre a realização do tratamento e os respectivos agentes de tratamento, observados os segredos comercial e industrial; [...]". BRASIL - Lei 13.709/18, de 14/08/18, art.º 6.º.

[219] O regulamento europeu disciplina a questão do consentimento em seu arts.ºs 7.º e 8.º. No Direito brasileiro, a diretriz encontra-se positivada no art.º 7.º, I, da Lei 13.709/18. BRASIL - Lei 13.709/18, de 14/08/18, art.º 7.º.

Embora endereçados, primariamente, aos consumidores em suas relações com empresas de fornecimento de produtos e serviços os referidos princípios podem assegurar a compatibilização entre privacidade e desenvolvimento tecnológicos.

Devem, portanto, nortear os parâmetros arquiteturais quanto à elaboração e implantação de plataformas de IA, sobretudo quando destinadas à potencialização de instrumentos tecno-normativos.

2.6. Liberdade de expressão e Liberdade política

Curiosamente, uma das ameaças à liberdade de expressão resulta do combate à produção e disseminação de notícias falsas ("*fake news*"[220]), assim como de conteúdo qualificado como "discurso de ódio", mediante o emprego de sofisticadas ferramentas para sua identificação e neutralização, por parte de entes privados e públicos[221].

Trata-se de imenso desafio a ser enfrentado sobretudo pelos órgãos legiferantes e jurisdicionais, na tentativa de promover o equilíbrio entre a tutela da verdade e a garantia de exercício da liberdade de expressão.

Os mecanismos de IA, nessa perspectiva, podem figurar tanto como impulsionadores quanto algozes da liberdade de expressão.

No primeiro angulo, os instrumentos tecnológicos (movidos por IA) que viabilizam as redes de comunicação têm assegurado uma plataforma aberta para exposição livre de ideias e o fluxo quase irrestrito de manifestações de toda ordem.

Essa mesma possibilidade oferece, contudo, as condições para a propagação de discursos radicais e repletos de informações falsas, cuja veracidade nem sempre pode ser verificada.

Quanto à liberdade política, em particular, encontrar-se-ia ameaçada em diversas frentes pela inteligência artificial, entre as quais se poderiam realçar, ao menos, dois fenômenos de alguma relevância e atualidade.

---

[220] Uma das preocupações relativas às "*fake news*" refere-se ao derruimento da "voz dos especialistas", assim como das instituições que promovem a "curadoria" do conhecimento humano, levando à erosão do próprio conceito de "dado objetivo", de forma a comprometer o engajamento no discurso racional, a partir de premissas fáticas compartilhadas. É o que aponta relevante estudo conduzido pela Universidade de Cambridge, disponível em: https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/abs/governing-fake-news-the-regulation-of-social-media-and-the-right-to-freedom-of-expression-in-the-era-of-emergency/21D0AD592E61726EEC999567ADFB8246

[221] ACCESS NOW – *Human Rights in the Age of Artificial Intelligence*, p. 22.

2.6.1. "*Fake news*"

O primeiro refere-se ao problema da propagação das denominadas "*fake news*", fortemente impulsionadas por ferramentas de IA, a polarizarem e reforçarem as cisões entre os cidadãos no seu exercício do direito à intervenção no processo eleitoral, por meio da intervenção nos fluxos de formação do convencimento político.

O emprego da IA atua não apenas na otimização dos mecanismos de difusão de tais notícias, mas também na própria criação de falsas informações a cada dia mais convincentes[222].

Nesse contexto, também pode ser referido o uso da tecnologia conhecida como "*deep fake*", que permite a adulteração e manipulação de vídeos com altíssimo nível de realismo. O seu potencial de uso como "arma política" tem se confirmado a partir de exemplos provindos de vários países[223].

Há de se ponderar, de outro lado, a possibilidade de utilização de sistemas de IA para identificar e combater a propagação de notícias falsas, em um contexto "*infodêmico*", i.e., de excesso informacional[224]. Dado o volume imenso de informações a circularem nos meios digitais[225], apenas soluções robustas poderiam dar cabo da ingente tarefa de monitorar o tráfego, à busca de padrões ou distorções que possam denunciar a circulação de tais rumores fabricados, além de encontrar a própria fonte da "*desinformação*" e auxiliar na realização um trabalho de checagem ou apuração da verdade ("*fact-finding*")[226].

Têm sido amplamente noticiados os impactos nefastos das sofisticadas formas de indução de comportamento coletivo que podem ser observadas nas redes sociais, que teriam

---

[222] A esse propósito: https://scroll.in/article/997898/fake-news-generated-by-artificial-intelligence-can-be-convincing-enough-to-trick-even-experts [Consult. 05 nov. 2021].

[223] "*In Gabon, the military launched an ultimately unsuccessful coup after the release of an apparently fake video of leader Ali Bongo suggested that the President was no longer healthy enough to hold office. In Malaysia, a video purporting to show the Economic Affairs Minister having sex has generated a considerable debate over whether the video was faked or not, which caused reputational damage for the Minister. In Belgium, a political group released a deepfake of the Belgian Prime Minister giving a speech that linked the COVID-19 outbreak to environmental damage and called for drastic action on climate change*". Fonte: https://www.weforum.org/agenda/2020/10/deepfake-democracy-could-modern-elections-fall-prey-to-fiction/ [Consult. 05 nov. 2021].

[224] Dentre as diversas fontes a esse respeito, pode ser consultada: https://www.forbes.com/sites/bernardmarr/2021/01/25/fake-news-is-rampant-here-is-how-artificial-intelligence-can-help/ [Consult. 05 nov. 2021].

[225] Estima-se que, a cada minuto, cerca de 160.000.000 de e-mails sejam enviados, 98.000 tweets sejam publicados, 600 vídeos sejam disponibilizados no Youtube. Fonte: https://bernardmarr.com/fake-news-and-how-artificial-intelligence-tools-can-help/ [Consult. 05 nov. 2021].

[226] Para uma análise mais técnica das possibilidades e limites das contribuições da IA nesse campo, pode ser recomendado o estudo de HORNE e outros: HORNE, Benjamin D. & NEVO, Dorit & ADALI, Sibel & MANIKONDA, Lydia & ARRINGTON, Clare – *Tailoring heuristics and timing AI interventions for supporting news veracity assessments*.

afetado até mesmo o destino de eleições mesmo em países com sólida tradição democrática[227].

2.6.2. Vigilância eletrônica

O segundo seria o resultante das limitações aos movimentos e manifestações populares, ante o recorrente uso de instrumentos de vigilância eletrônica, v.g., câmeras com reconhecimento facial[228].

As características dessa supervisão agravam as preocupações em torno das suas implicações quanto às liberdades individuais, com destaque para a liberdade política: i) ubiquidade ou capilaridade – os pontos de obtenção de dados estão presentes em quase todas as dimensões da vida cotidiana das pessoas, especialmente ante a forte presença digital[229]; ii) invasividade – um espectro abrangente das informações comportamentais tem sido registrado por entidades privadas e públicas; iii) caráter "panóptico" – por não se saber quando e o que está sendo objeto de vigilância, por vezes.

Nesse ponto, instala-se o debate entre a garantia de segurança e o possível comprometimento das liberdades de movimento além, naturalmente, da já mencionada privacidade.

2.7. Trabalho

O senso comum sugere uma quase inexorável correlação entre evolução tecnológica e desemprego ou degradação das condições de trabalho, sobretudo em atividades repetitivas, o que, contudo, tem sido objeto de questionamentos no plano da investigação empírica.

---

[227] Emblemático o conhecido escândalo da empresa Cambridge Analytica que teria levado à imposição de expressiva multa (cinco mil milhões de euros) ao Facebook, por violação de regras concernentes à proteção de dados pessoais. A coleta de informações por parte daquela empresa (Cambridge Analytica) teria permitido a criação de perfis falsos destinados à propagação de mensagens políticas. Entre outras fontes, pode ser consultada a matéria divulgada pelo Diário de Notícias em: https://www.dn.pt/vida-e-futuro/facebook-multado-em-cinco-mil-milhoes-de-dolares-pelo-escandalo-cambridge-analytica-11108647.html [Consult. 05 nov. 2021].

[228] Hong Kong representa um importante exemplo dos riscos de comprometimento das liberdades políticas, a partir do uso de tecnologias invasivas, com forte impulsionamento da IA. A identificação e detenção dos participantes dos protestos ocorridos em 2020 deu-se graças ao cruzamento de informações obtidas a partir do monitoramento dos sistemas de transporte público (dados dos cartões de mobilidade) e das câmeras de segurança, espalhadas pela cidade. A esse respeito: https://www.nytimes.com/2020/06/12/technology/surveillance-protests-hong-kong.html [Consult. 05 nov. 2021]. Para um ranking das cidades mais vigiadas do mundo, a partir do número de câmeras por habitante: https://sg.news.yahoo.com/singapore-11th-most-surveilled-city-062420700.html [Consult. 05 nov. 2021].

[229] Para um ranking das cidades mais vigiadas do mundo, a partir do número de câmeras por habitante: https://sg.news.yahoo.com/singapore-11th-most-surveilled-city-062420700.html [Consult. 05 nov. 2021].

No caso da inteligência artificial, há um forte temor de que a sua adoção represente, inevitavelmente, a supressão de postos de trabalho, mesmo no caso de labor com alguma complexidade cognitiva e sem recorrência.

Entretanto, alguns estudos indicam a ausência de comprovação de um declínio inevitável e generalizado dos empregos em ocupações mais suscetíveis à exposição às novas tecnologias movidas pela IA[230].

Seja como for, parece indene de dúvidas a existência de profundos impactos resultantes de tais tecnologias em todas as fases de desenvolvimento dos contratos de trabalho, desde a sua formação até a eventual extinção.

2.7.1. Acesso ao emprego

Duas relevantes dimensões podem ser consideradas ao se abordar a questão do impacto da IA quanto ao acesso ao emprego: i) a diminuição (ou deslocação) de postos de trabalho como resultado de eventual automação; e ii) a mudança no processo de recrutamento e seleção de pessoal, com apoio em ferramentas artificialmente inteligentes.

Em relação à primeira perspectiva, há uma grande diversidade de previsões, com alguns pontos convergentes.

Um dos mais evidentes concerne à natural vantagem dos países desenvolvidos, com alta renda média, sobretudo entre os que já contam com um nível maior de tecnologia incorporada no cotidiano, quanto aos quais se descortina um horizonte favorável às adaptações necessárias à preservação e mesmo incremento dos empregos em um ambiente de grande automação movida pela IA[231].

Outra quase certeza diz respeito à marcada tendência de ampla eliminação das ocupações baseadas em rotinas bem definidas ou dependentes de atividades estritamente

---

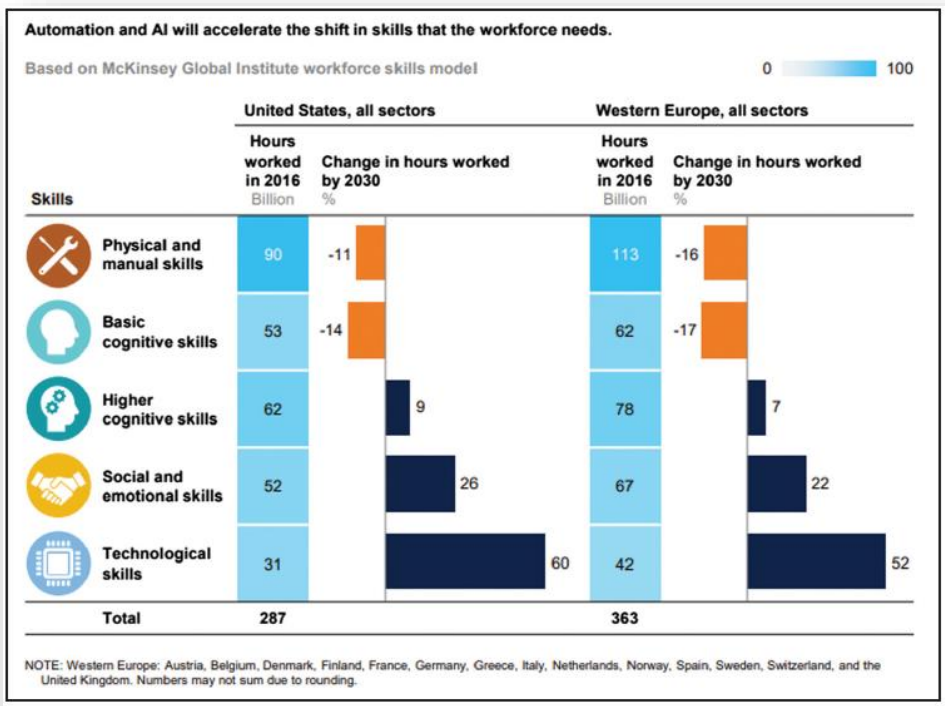
[230] "*From a theoretical perspective, the impact of AI on employment and wages is ambiguous, and it may depend strongly on the type of AI being developed and deployed, how it is developed and deployed, and on market conditions and policy. However, the empirical evidence based on AI adopted in the last 10 years does not support the idea of an overall decline in employment and wages in occupations exposed to AI*". OECD – *The impact of Artificial Intelligence on the labour market: What do we know so far?*.

[231] "*Thanks to the availability of robust, long-term employment and labour market data, more evidence on the impact of AI is available for high-income countries, which include Organisation for Economic Co-operation and Development (OECD) countries and many EU Member States. Cross-analysis of forecasts, including those provided by OECD and McKinsey & Company on AI and job losses, suggests that up to one third of work activities could be displaced by 2030 due to automation. High-income economies are expected to be better positioned to adapt to the changes required by AI-led automation than middle- to low-income ones*" *EUROPEAN PARLIAMENT'S SPECIAL COMMITTEE ON ARTIFICIAL INTELLIGENCE IN A DIGITAL AGE (AIDA) – Improving working conditions using Artificial Intelligence*, p. 11.

manuais e repetitivas[232]. Trata-se do resultado da provável aceleração da mudança de distribuição das habilidades laborativas demandadas no mercado de trabalho, a respeito da qual há diversas projeções, entre as quais se pode destacar a seguinte[233]:

Figura 6.

Mudança de distribuição das habilidades laborativas demandadas no mercado de trabalho[234]



[232] "*Occupations that rely on well-defined routines and occupations involving physical work/manual work that rely solely on physical strength are at the most risk of loss of demand. This includes jobs in agriculture (due to the development of remotely controlled, self-driving tractors); in call centres (which can be replaced by speech recognition software); in postal services (mail sorters, processors, carriers); in clerical work (data entry, legal); and in retail (shop assistants). Jobs in unpredictable environments (such as gardening, plumbing, childcare and care of the elderly) are considered technically difficult to automate and so are projected to be at lower risk of automation by 2030. Occupations involving non-routine cognitive tasks (such as lab technicians, engineers and actuaries) are often judged to be most exposed to AI. Forecasts highlight that jobs requiring social and cognitive intelligence and empathy, as well as jobs in unpredictable environments, are expected to be at low risk from AI*". *EUROPEAN PARLIAMENT'S SPECIAL COMMITTEE ON ARTIFICIAL INTELLIGENCE IN A DIGITAL AGE (AIDA) – Op. cit.*, p. 12.

[233] ERNST, Ekkehardt & MEROLA, Rossana & SAMAAN, Daniel – *Economics of artificial intelligence: Implications for the future of work*, p. 21.

[234] ERNST, Ekkehardt & MEROLA, Rossana & SAMAAN, Daniel – *Op. cit.*, p. 21.

Constata-se, de outro lado, uma possibilidade de deslocamento dos empregos, graças à mudança na divisão do trabalho entre homens e máquinas[235]. Haveria perdas em determinados segmentos e ganhos em outros[236], em um contexto de migração já em curso da manufatura para os serviços[237], além de uma profunda alteração em diversas tarefas, sem que isso venha a levar necessariamente à supressão de postos de trabalho[238].

No tocante à mudança no processo de recrutamento e seleção de pessoal, há avaliações antagônicas a respeito dos riscos derivados da aplicação da IA.

Há, de um lado indicações no sentido de que o processo seletivo pode se beneficiar de uma análise mais objetiva, lastreada em indicadores extraídos de sistemas artificialmente inteligentes[239].

De outro, existem estudos a apontarem sensível enviezamento algorítmico em desfavor de algumas minorias (étnicas e culturais) ou grupos particularmente vulneráveis (em razão do gênero ou do extrato social)[240].

2.7.2. Execução dos contratos laborais

Na execução dos contratos de trabalho, podem ser identificadas algumas preocupações em torno do uso da IA em algumas dimensões relevantes dos direitos dos

---

[235] "*The WEF 2020 Future of Jobs survey respondents estimated that by 2025, 85 million jobs may be displaced by a shift in the division of labour between humans and machines*". *EUROPEAN PARLIAMENT'S SPECIAL COMMITTEE ON ARTIFICIAL INTELLIGENCE IN A DIGITAL AGE (AIDA) – Op. cit.*, p. 13.

[236] "*The banking, transport and health sectors are identified as examples of sectors which have adapted to such a transformation by modifying job requirements, reskilling workers and developing new services. The agricultural sector faces job losses in high-income countries. Due to the use of automatic irrigation systems and intelligent harvesting and transporting machines, the demand for seasonal, temporary labour (which is often employed through foreign harvest hands and helpers) is expected to decrease*". *EUROPEAN PARLIAMENT'S SPECIAL COMMITTEE ON ARTIFICIAL INTELLIGENCE IN A DIGITAL AGE (AIDA) – Op. cit.*, p. 14.

[237] SERVOZ, Michel - *THE FUTURE OF WORK? THE WORK OF THE FUTURE!*, p. 12.

[238] "*Although there have been notable forecasts identifying large-scale disruption of current job markets, AI is expected to displace parts of the activities within jobs over the long term, rather than replace entire jobs. Very few occupations consist primarily of performing automatable tasks, although nearly all occupations include automatable components. This view of AI and robotics displacing tasks rather than jobs in their entirety was also echoed by some interviewees*". *EUROPEAN PARLIAMENT'S SPECIAL COMMITTEE ON ARTIFICIAL INTELLIGENCE IN A DIGITAL AGE (AIDA) – Op. cit.*, p. 15.

[239] "*Some sources consider that algorithms, which use AI to make decisions based on existing datasets, may provide more objective and neutral ways of measuring employees' performance and eliminate the possibility of individual biases by supervisors […]. Similarly, in recruitment decisions, there is potential for tools using AI to reduce bias against groups of applicants, again ensuring more inclusion and diversity in the workplace. Other sources, however, identified more risks than benefits in this regard". EUROPEAN PARLIAMENT'S SPECIAL COMMITTEE ON ARTIFICIAL INTELLIGENCE IN A DIGITAL AGE (AIDA) – Op. cit.*, p. 19.

[240] Um ilustrativo exemplo foi o que se deu com a empresa Amazon, a qual teria se valido de ferramenta de recrutamento dotada de algoritmo com notável viés prejudicial às mulheres. Entre outras fontes, pode ser mencionada: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G [Consult. 14 jun. 20212].

trabalhadores, entre as quais: i) integridade física e psíquica; ii) privacidade; iii) isonomia; e iv) liberdade associativa.

Relativamente à integridade física, os riscos concernem, exemplificativamente, às possíveis falhas operacionais em máquinas operadas, total ou parcialmente, por agentes artificialmente inteligentes[241].

O comprometimento psíquico dos trabalhadores poderia decorrer do desenvolvimento de transtornos resultantes, entre outros fatores, do pavor da iminente substituição por máquinas, do convívio como os robôs[242] (nas mais diferentes acepções do termo[243]), assim como da diminuição ou mesmo eliminação da interação presencial com outras pessoas[244].

Também poderia contribuir para a instabilização da vigilância panóptica a que estariam sujeitos por meio de câmeras, microfones, softwares de monitoramento de desempenho, relógios inteligentes que captam informações sobre o estado de saúde[245].

Além dos suprareferidos instrumentos de vigilância, há outras fontes de vulnerabilidade ao direito de privacidade no ambiente laboral. Representa uma importante ameaça à proteção das informações pessoais dos trabalhadores a utilização de sistemas de IA para a realização da coleta, armazenamento e manuseio de dados obtidos de forma massiva ("big data"), automatizada e, por vezes, sub-reptícia, por parte dos empregadores[246].

---

[241] Um dos exemplos citados pela Comissão Europeia diz respeito à coleta mecanizada de frutos que em atividades agrárias, nas quais o uso de tecnologia assistida por IA produziria risco de lesão física, em caso de falha operacional. Eis o que afirma em relatoria a Comissão: "Embora os algoritmos desenvolvidos já apresentem taxas de sucesso na classificação superiores a 90 %, uma deficiência nos conjuntos de dados que alimentam esses algoritmos pode levar os robôs a tomarem más decisões e, consequentemente, a ferirem animais ou pessoas". COMISSÃO EUROPEIA - Relatório sobre as implicações em matéria de segurança e de responsabilidade decorrentes da inteligência artificial, da Internet das coisas e da robótica, p. 10.

[242] "*Workers' tasks may also become increasingly fragmented when AI is used to take over some or all aspects. This may result in workers feeling that they do not understand the entire task, feeling less interested, or feeling that they are being de-skilled and under-valued*". *EUROPEAN PARLIAMENT'S SPECIAL COMMITTEE ON ARTIFICIAL INTELLIGENCE IN A DIGITAL AGE (AIDA) – Op. cit.*, p. 20.

[243] Aqui se alude tanto aos robôs humanoides, como aos braços mecânicos, máquinas sofisticadas em linhas de produção ou mesmo programas de computador que realizem atividades com algum nível de autonomia.

[244] "*The European Parliament discussed the risks of dehumanisation within those receiving services from AI technologies, but fewer studies have discussed the similar pressure on workers collaborating with automated technology. Workers who increasingly work alongside AI-powered algorithms instead of colleagues or managers may experience a decline in their social interaction. Relationships with colleagues and co-workers are considered to be important for workers' motivation and wellbeing, and the elimination of such interactions from workplaces may lead to dissatisfaction and less motivation*". *EUROPEAN PARLIAMENT'S SPECIAL COMMITTEE ON ARTIFICIAL INTELLIGENCE IN A DIGITAL AGE (AIDA) – Op. cit.*, p. 20.

[245] MANTELLO, Peter & HO, Manh-Tung & NGUYEN, Minh-Hoang & VUONG, Quan-Hoang - *Bosses without a heart: socio-demographic and cross-cultural determinants of attitude toward Emotional AI in the workplace*.

[246] IBA GLOBAL EMPLOYMENT INSTITUTE - *Artificial Intelligence and Robotics and Their Impact on the Workplace*, p. 102.

A isonomia pode vir a ser afetada no contexto das relações de trabalho a partir de diferentes perspectivas, entre as quais pode ser lembrada a diferenciação salarial fixada a partir de avaliações algorítmicas de performance, contaminadas por possíveis enviezamentos em desfavor de minorias ou grupos mais vulneráveis[247]. De outro lado, a própria dificuldade de absorção das tecnologias ligadas à IA[248] pode levar à ampliação da desigualdade remuneratória entre trabalhadores.

Quanto à liberdade associativa, uma das inquietações repousa na possível coletada por ferramentas de IA de dados dos trabalhadores (v.g. localização, teor de mensagens de email, captação de imagem e/ou áudio) para identificar eventuais participações em reuniões para conduzir movimentos reivindicatórios ou mesmo para fundação de entidades sindicais, o que também se relaciona com a questão da privacidade antes referida[249].

Além disso, há algum temor quanto ao uso de algoritmos[250] nas negociações coletivas, com a ampliação das vantagens negociais dos empregadores[251], além de possível frustação do direito de acesso a informações vitais ao processo negocial, dada a opacidade dos sistemas. Cumpre observar que um dos caminhos aventados para conter tais riscos seria o fortalecimento dos entes coletivos[252].

---

[247] "*There is also a risk that AI may perpetuate bias due to having been developed inlargely homogenous environments (including, for example, a high proportion of white men). To mitigate these risks, the data used needs to be as unbiased as possible, and also varied, to ensure that AI software can make balanced decisions*". *EUROPEAN PARLIAMENT'S SPECIAL COMMITTEE ON ARTIFICIAL INTELLIGENCE IN A DIGITAL AGE (AIDA) – Op. cit.*, p. 20.

[248] Por motivos de ordem social, cultural ou econômica.

[249] Caso rumoroso, nesse sentido, foi o da denúncia de que empresas como a AMAZON estariam a coletar informações das redes sociais de seus empregados para monitorar o seu "ativismo" sindical. Fonte: https://www.npr.org/2020/11/30/940196997/amazon-reportedly-has-pinkerton-agents-surveil-workers-who-try-to-form-unions?t=1656100851723 [Consult. 26 jun. 2022]

[250] Trata-se de um dos resultados de uma abordagem que alguns chamam de "*algorithmic management*" or "*management-by-algorithm*". DE STEFANO, Valerio e TAES, Simon - "*Algorithmic management fits into a broader context of the implementation of technological tools and digitalised supervision systems aimed at governing the workforce*", p. 3.

[251] *EUROPEAN PARLIAMENT'S SPECIAL COMMITTEE ON ARTIFICIAL INTELLIGENCE IN A DIGITAL AGE (*AIDA) – Op. cit., p. 23.

[252] Eis o que assenta a Organização Internacional do Trabalho a esse respeito: "*Collective agreements could address the use of digital technology, data collection and algorithms that direct and discipline the workforce, ensuring transparency, social sustainability and compliance with these practices with regulation. Collective negotiation would also prove pivotal in implementing the "human-in-command" approach at the workplace. Collective bargaining could also regulate issues such as the ownership of the data collected from workers and go as far as creating bilateral or independent bodies that would own and manage some of the data*" *INTERNATIONAL LABOUR ORGANIZATION – EMPLOYMENT POLICY DEPARTMENT* – STEFANO, Valerio de – *"Negotiating the algorithm": Automation, artificial intelligence and labour protection*, p. 23.

2.7.3. Extinção dos contratos laborais

No capítulo relativo ao encerramento das relações jurídico laborais pode-se divisar variados impactos do uso da IA na seara laboral.

Em primeiro lugar, o desligamento pode ser o resultado do já abordado processo de automação oriundo da implantação de sistemas de IA[253]. Nesse caso, o agravante consiste na tendência de promoção de dispensas em escala, a afetar coletividades de trabalhadores, com indesejáveis danos econômicos e sociais[254].

Nesse ponto, há importantes marcos normativos a traçarem limites ao exercício do direito potestativo dos empregadores, de forma a conter os consectários deletérios de tais rupturas em massa. No âmbito internacional, a referência mais relevante seria a Convenção 158 da Organização Internacional do Trabalho, que prevê requisitos formais e materiais para concretização dos desligamentos coletivos[255]. Já na seara europeia, pode ser mencionada a Diretiva 1998/59.

De outra parte, a própria decisão pela terminação do contrato de trabalho pode derivar de uma análise realizada a partir de instrumentos tecnológicos dotados de IA, o que pode, mais uma vez, trazer um notável risco de quebra de isonomia[256].

Ademais, como resultado direto ou indireto da penetração das tecnologias associadas à IA, a fragmentação dos empregos em tarefas poderá levar a uma diminuição da estabilidade contratual[257]. Daí decorreria a preferência por contratos a termo, focados em atividades específicas e limitadas no tempo[258].

---

[253] No âmbito europeu, entre outros estudos, merece destaque o multicitado relatório do Comitê *AIDA. EUROPEAN PARLIAMENT'S SPECIAL COMMITTEE ON ARTIFICIAL INTELLIGENCE IN A DIGITAL AGE* (AIDA) – *Op. cit.*

[254] De acordo com relatório da Organização Internacional de Trabalho (OIT), apresentado em 2016, o motivo mais recorrente para dispensas coletivas seria a reestruturação interna das empresas, o que, em boa parte dos casos, pode derivar de processos de automação. No mesmo texto, a OIT também apresenta uma relação de efeitos de tais desligamentos em massa, como seria o caso também de consequências quanto à própria saúde, física e psicológica, dos trabalhadores afetados. *INTERNATIONAL LABOUR OFFICE – Report on collective dismissals - A comparative and contextual analysis of the law on collective redundancies in 13 European countries*.

[255] Na Parte III do instrumento, trata-se da terminação decorrente de razões de ordem econômica, tecnológica ou estrutural.

[256] Notícia pertinente a esse respeito foi divulgada por diversos meios de comunicação quanto ao desligamento por parte da empresa russa XSOLLA de 150 de seus empregados (1/3 da sua força de trabalho global, composta de 450 trabalhadores), a partir de decisão tomada por algoritmos. Fonte: https://english.elpais.com/usa/2021-10-14/one-second-150-dismissals-inside-the-algorithms-that-decide-who-should-lose-their-job.html [Consult. 29 jun. 2022].

[257] *EUROFOUND – Game-changing technologies: Transforming production and employment in Europe*, p. 37.

[258] *EUROFOUND – Op. cit.*, p. 37.

3. Regulação atual

3.1. Modelos regulatórios

A regulação da tecnologia representa um especial desafio às instâncias legiferantes estatais, ante suas características intrínsecas, a ponto de alguns a considerarem inviável, ao menos em algumas das suas dimensões.

Tal afirmação decorre, em primeiro lugar, da própria dificuldade de compreensão em nível até mesmo conceitual do objeto a ser regulado, dada a eventual complexidade técnica ou científica do tema (v.g. em situações como células-tronco, nanotecnologia e clonagem).

Além disso, a ausência de agilidade nos processos legislativos em países democráticos parece não condizer com o intenso ritmo evolutivo das transformações tecnológicas.

Aqui repousam dois obstáculos particulares: i) o que alguns denominam "*pacing problem*"[259] (problema do "ritmo", em uma tradução literal); e ii) o chamado "paradoxo" ou "dilema de Collingridge"[260].

---

[259] Na feliz síntese de MARCHANT: "*The central conclusion from the cumulative insights of the various contributions […] is that existing regulatory systems and ethical frameworks are inadequate to provide effective, meaningful and timely oversight of the current and futuregenerations of emerging technologies. The GRINN technologies – genetics, robotics, information technologies, nanotechnology, and neuroscience – are racing ahead at a pace of technology development that has never before been experienced in human history. In contrast, our traditional government oversight systems are mired in stagnation, ossification and bureaucratic inertia, and are seriously and increasingly lagging behind the new technologies accelerating into the future*". MARCHANT, E. Gary – *Addressing the Pacing Problem*, p. 199.

[260] Assim o descreve MOSES: "*One aspect of newness that is important from a regulatory perspective is the so-called Collingridge dilemma. Collingridge was concerned that regulators responding to a new technology faced twin hurdles. At an early stage in a technology's development, regulation was problematic due to the lack of information about the technology's likely impact. At a later stage, regulation was problematic as the technology would become more entrenched, making any changes demanded by regulators expensive to implement. The dilemma follows from sociological studies of technology that suggest that 'interpretive flexibility' is high in the early stages of a technology's development, but ultimately stabilises in a (more or less) final form ('closure') following stabilisation. Another way of understanding the same phenomenon is to recognise that technological systems acquire 'momentum' as they grow larger and more complex, making them more resistant to regulatory prodding. This suggests that regulators wishing to influence technological design (to avoid or minimise risks of health, environmental and social harm, for instance) need to act at an early stage when the situation is more malleable. At an early stage, however, little is known about the prospects for the new technology, the harms it might cause or the forms it might take. Thus regulators face an 'uncertainty paradox',42 where they are forced to make decisions in the absence of reliable risk information or foreknowledge of technological developments. The extent to which these twin obstacles prove to be a dilemma depends on the rapidity and unpredictability of technological change, as well as the diffusion pattern associated with the technology in question. There may be partial solutions to this problem, such as involving experts, improving understandings of how regulators can manage different types of uncertainty, expressing obligations in broad terms or adopting a particular approach (such as the precautionary principle). Evaluating such solutions, alone or in combination, is important as it could guide government decision-making processes in new technological contexts*". MOSES - Lyria Bennett - *How to Think about Law, Regulation and Technology: Problems with 'Technology' as a Regulatory Target*, pp. 7-8.

O primeiro ("*pacing*") consiste, essencialmente, no descompasso quase inevitável entre avanços tecnológicos e regulação. Resulta de uma miríade de causas, para além da evidente assimetria na cadência evolutiva entre os domínios tecnológico e jurídico[261], a exigir o enfrentamento a partir de diferentes perspectivas[262].

MARCHANT indica, essencialmente, três caminhos a serem trilhados para enfrentar o problema de "pacing": i) governança adaptativa ("*adaptive governance*") – processo interativo que envolveria continuas atividades de monitoramento, avaliação e ajuste da supervisão regulatória; ii) mecanismos de "*soft law*"; e iii) reformas institucionais[263].

O segundo (paradoxo de "*Collingridge*") está relacionado com a dificílima tarefa de identificar o melhor momento para promover a intervenção regulatória em uma trajetória temporal não necessariamente linear na qual seriam sopesados o conhecimento sobre a matéria (as informações disponíveis de início seriam escassas), a adesão à tecnologia pela sociedade (sua incorporação massiva pode tornar ineficazes as tentativas de regular o fenômeno) e seus efeitos potencialmente lesivos aos direitos e garantias individuais e coletivos[264].

Ante a ubiquidade própria à disseminação de tecnologias especialmente as relativas à informatização, de grande capacidade de propagação (v.g. aplicações de telemóveis e programas de computador), outro ponto a ser considerado é a dificuldade em se reunir consenso de forma suficientemente abrangente para estabelecer marcos regulatórios em escala global ou transfronteiriça.

A liberdade científica e a preocupação com o possível estrangulamento da inovação[265] também atuam como restrições importantes aos instrumentos de regulação pela via estatal.

---

[261] "*The challenge to law and ethics to keep pace with rapidly developing emerging technologies is affected by other dynamics in addition to the disparity of the relative speeds of the two domains. The oversight of emerging technologies is more complex than most previous regulatory challenges, in that the technologies involved tend to have many diverse applications and forms, are used in many different industries and contexts, and present a multitude of different and often hard-to-quantify risk and benefit scenarios. These technologies often fail to fit comfortably within existing regulatory categories, and thus the path dependency created by regulatory frameworks developed for earlier, simpler technologies can be problematic*". MARCHANT, E. Gary – *Op. cit.*, p. 199.

[262] MARCHANT, E. Gary – *Op. cit.*, p. 201.

[263] MARCHANT, E. Gary – *Id. ibid.*

[264] MOSES - Lyria Bennett - *How to Think about Law, Regulation and Technology: Problems with 'Technology' as a Regulatory Target*, pp. 7-8.

[265] A respeito da inovação, podem ser citadas algumas das interessantes conclusões de relatório publicado pela Organização Mundidal da Propriedade Intelectual (*WIPO - World Intellectual Property Organization*), entre as quais a de que mais da metade de todas as patentes a respeito de IA foi registrada a partir de 2013: "*Since artificial intelligence emerged in the 1950s, innovators and researchers have filed applications for nearly 340,000 AI-related inventions and published over 1.6 million scientific publications. Notably, AI-related patenting is growing rapidly: over half of the identified inventions have been published since 2013. While scientific publications on AI date back decades, the boom in scientific publications on AI only started around 2001, approximately 12 years in advance of an upsurge in patent applications. Moreover, the ratio of scientific papers to inventions has decreased from 8:1 in 2010 to 3:1 in 2016 – indicative of a shift from theoretical research to the use of AI technologies in*

Tais aspectos indicariam uma certa tendência de primazia de soluções autorregulatórias[266].

Durante algum tempo, ilustrativamente, não se entendia "possível" nem "desejável" a imposição de normas jurídicas a respeito da rede mundial de computadores. A internet era vista como um "território livre"[267], onde as amarras estatais não tinham lugar e a conduta humana não deveria encontrar limites jurídicos (ou morais) apriorísticos.

O célere desenvolvimento da internet, com a implementação de ferramentas de interação e acesso a informações em escala nunca antes vista fez emergir forte convicção acerca da conveniência e necessidade, em alguns casos, de definição de marcos jurídicos[268]. Os incontestes impactos sociais, econômicos e jurídicos tornaram premente a intervenção regulatória do Estado e de outros entes com capacidade de produção normativa (v.g. organizações internacionais).

O mesmo pode ser afirmado a respeito da Inteligência Artificial, fenómeno que talvez possa ser considerado ainda em sua "infância" e quanto ao qual, para além das questões já mencionadas, podem ser identificados problemas específicos, entre os quais se poderia indicar: i) "*discreetness*" – desenvolvimento sem uma estrutura institucional integrada e de larga escala; ii) "*difuseness*" – existência de difusão de atores e contextos (geográficos) heterogêneos; iii) "*discretness*" – o uso de um mosaico de elementos tecnológicos "ocultos" a obstar a identificação dos riscos até o momento a implementação dos sistemas; iv) "*opacity*" – a opacidade sistêmica de boa parte dos projetos a impedir a supervisão e auditagem dos sistemas; v) "*foresseability*" – a imprevisibilidade em diferentes níveis resultante dos graus de autonomia presentes nos agentes artificialmente inteligentes a produzir lacunas na imputação de responsabilidade por eventuais danos; vi) "*narrow control*" – o uso dos sistemas de IA à revelia daqueles que seriam legalmente responsáveis por suas

commercial products and services". *WORLD INTELLECTUAL PROPERTY ORGANIZATION - WIPO Technology Trends 2019 Executive summary Artificial Intelligence*, p. 6.

[266] "*The dynamics of scientific research, requirements for autonomy and flexibility, ethical considerations brought by highly sensitive domains, and above all the principle of the freedom of research account for the importance lent to self-regulation as the logical regulatory framework from very early on by the researchers themselves*". GONÇALVES, Maria Eduarda & GAMEIRO, Maria – *Hard Law, Soft Law and Self-regulation: Seeking Better Governance for Science and Technology in the EU*, p. 3.

[267] Emblemática, a propósito, a conhecida Declaração de Independência do Cyberespaço de John Perry Barlow, publicada em 1996.

[268] No Brasil, apenas em 2014 é que se viu surgir um estatuto mais abrangente acerca da matéria. Trata-se da Lei n.º 12.965/14. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm [Consult. 19 jul. 2022].

"ações"; e vii) "*general control*" – ao alcançar a AGI ou a SGI, os agentes se desgarrariam por completo do "controle humano"[269].

Alguns desses desafios se colocariam "*ex ante*", i.e., durante a pesquisa e o desenvolvimento dos sistemas de IA. Já outros seriam "*ex post*", i.e., apenas ocorreriam com a sua criação e implementação[270].

Nesse cenário, contemplar as tentativas regulatórias passadas e presentes na experiência comparada poder oferecer alguns "*insights*" quanto aos caminhos a serem explorados ou descartados, mas sobretudo deve reforçar a percepção quanto à imprescindibilidade e relevância de conceber ferramentas regulatórias eficientes, de modo a preservar a incolumidade sobretudo de direitos fundamentais.

Entre as opções regulatórias mais comumente lembradas, podem ser citados as seguintes: i) *hard law*; ii) *soft law*; iii) autoregulação ("*self-regulation*"); iv) autoregulação regulada; e v) tecnoregulação.

Trata-se de abordagens cujas fronteiras conceituais não podem ser tidas como estanques, redigidas ou inflexíveis. Antes, admitem alguma fluidez, reconhecendo-se a forte interação e interpenetração entre tais vias de regulação.

Ilustrativamente, os mecanismos de "*soft law*" e autoregulação atuam, por vezes, como preparatórios para soluções de "*hard law*". De outro lado, instrumentos normativos estatais podem positivar princípios ou regras gerais a serem observadas no plano autoregulatório.

Seja como for, já se pode antecipar alguma inclinação no sentido de se reconhecer a predominância atual de vias regulatórias alternativas[271] às tradicionais para enfrentar os intricados desafios jurídicos trazidos pela IA.

3.1.1. "Hard law"

Uma das maiores e mais óbvias deficiências da abordagem regulatória centrada em soluções de "s*oft law*" consiste na ausência de mecanismos de coerção tradicionais. Segundo

[269] SCHERER, Matthew U. – *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*. Harvard Journal of Law & Technology Volume 29, Number 2 Spring 2016, p. 359.
[270] SCHERER, Matthew U. – *Op. cit.*, p. 359.
[271] Na verdade, cuida-se de constatação aplicáveis às questões de ciência e tecnologia de forma geral. "[…] *there is a wide belief that soft law, together with voluntary self-regulation, provide suitable regulatory tools for S&T* [*Science and Technology*]*, possibly better tools than 'hard law', to cope with the need for both flexibility and adjustment to novelty and prevailing uncertainties*". GONÇALVES, Maria Eduarda & GAMEIRO, Maria – *Hard Law, Soft Law and Self-regulation: Seeking Better Governance for Science and Technology in the EU*, p. 3.

levantamento realizado pela Universidade do Arizona, cerca de 70% dos instrumentos não dispõem de qualquer ferramenta coercitiva[272].

Haveria, nesse contexto, alguma tendência no Direito Comparado em se voltar para a "*hard law*", diante dos riscos já descortinados até aqui, inclusive quanto a direitos fundamentais universalmente reconhecidos (v.g. vida e integridade física).

Trata-se do que determinados autores vislumbram como um terceiro estágio dentro de um processo evolutivo composto das seguintes fases: i) divulgação de diretrizes principiológicas e éticas ("*AI ethics guidelines*") por parte de empresas, instituição acadêmicas e entidades governamentais, caracterizadas por serem genéricas e pouco concretas; ii) publicação de diversos textos científicos a revelarem algum consenso quanto à insuficiência de mecanismos de "soft law" para resguardar garantias fundamentais; e iii) desenvolvimento de normas legais dotadas de força coercitiva[273].

Seja como for, até o presente momento, não se identifica qualquer legislação nacional[274] em vigor a regular de forma abrangente[275] por meio de "*hard law*" os diferentes aspectos da inteligência artificial e seus impactos nos direitos e interesses juridicamente relevantes[276]. Mais adiante serão abordadas algumas tentativas em curso no âmbito europeu e no Direito brasileiro.

Por outro lado, há uma pletora diversificada de instrumentos regulatórios de *soft law* sobre IA, dada a amplitude conceitual do termo, a contemplar desde "*standards*" privados, "melhores práticas", códigos de conduta, declarações de princípios, até guias gerais[277].

---

[272] *Arizona State University – Soft-Law Governance of Artificial Intelligence, Executive Sumary*.
[273] THILO, Hagendorff,– *The Ethics of AI Ethics: An Evaluation of Guidelines*.
[274] Há, contudo, notícias recentes de uma proliferação de projetos de lei em diversos países, entre os quais o Brasil, no qual podem ser referidos os seguintes, entre outros: i) PL 21/2020 (Câmara dos Deputados) - estabelece fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial no Brasil; ii) PL 240/2002 (Câmara dos Deputados) - cria a Lei da Inteligência Artificial, e dá outras providências; iii) PL 1969/2021 (Câmara dos Deputados - arquivado) - dispõe sobre os princípios, direitos e obrigações na utilização de sistemas de inteligência artificial; iv) PL 872/2021 (Senado Federal) - dispõe sobre o uso da Inteligência Artificial; v) PL 5691/2019 (Senado Federal) - Institui a Política Nacional de Inteligência Artificial; vi) PL 5051/2019 (Senado Federal) - Estabelece os princípios para o uso da Inteligência Artificial no Brasil. Fontes: https://www.camara.leg.br/busca-portal/proposicoes/pesquisa-simplificada e https://www25.senado.leg.br/web/atividade [Consult. 12 ago. 2022].
[275] No entanto, já se encontram em vigor vários atos normativos a disciplinarem aspectos pontuais ou específicos da IA, como é o caso daqueles noticiados *NCSL* (*National Conference of State Legislatures*) nos Estados Unidos da América, em que apenas no ano de 2022 17 Estados contariam com regramentos acerca da matéria. Fonte: https://www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to-artificial-intelligence.aspx [Consult. 12 ago. 2022].
[276] Organização das Nações Unidas (ONU) - *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, p. 16.
[277] MARCHANT, Gary - *"Soft Law" Governance of Artificial Intelligence*, p. 12.

São oriundos de iniciativas de entes públicos e privados, de diferentes setores, com atores nacionais, internacionais e organizações não governamentais[278]. Em seguida passa-se à análise de alguns dos mais representativos desses instrumentos.

3.1.2. "*Soft law*"[279]

De acordo com pesquisa realizada pela Universidade do Arizona, houve a publicação de 634 instrumentos de "*soft law*" acerca de IA, entre os anos de 2001 e 2018. Destacou-se a circunstância de que cerca de 90% desse acervo normativo emergiram a partir de 2017, o que revela uma explosão regulatória razoavelmente recente. Ademais, identificou-se grande concentração geográfica, com a produção localizada, predominantemente, nos Estados Unidos e na Europa[280].

Em um breve recorte histórico, uma recorrente alusão entre os estudiosos quanto aos instrumentos reputados precursores da tentativa de estabelecer fronteiras ético-normativas quanto ao desenvolvimento de agentes artificialmente inteligentes é o conjunto de leis da robótica que emerge do clássico literário de ficção científica, publicado por ISAAC AZIMOV em 1942[281]. As três leis enunciadas por AZIMOV[282] revelam-se aplicáveis, em grande medida, no terreno da IA, sobretudo em face da preocupação com o desalinhamento de metas entre os agentes artificialmente inteligentes e os valores humanos fundamentais[283].

---

[278] Para uma relação bastante ampla, veja-se o exaustivo estudo conduzido pela OCDE, no qual são indicadas dezenas de instrumentos, além de uma análise panorâmica do status da questão regulatória da IA em diversos países. OCDE – *Artificial Intelligence in Society*.

[279] A maior parte das considerações contidas nesse item ("3.1.2. *Softlaw*") foi extraída de: PAULA, Gáudio Ribeiro de – A Recomendação da OCDE Sobre Inteligência Artificial e sua Relevância Como Instrumento de "*Soft Law*" para Contenção dos Riscos aos Direitos Fundamentais. Artigo apresentado como exigência parcial para obtenção do Grau de Doutoramento em Direito na disciplina "Estado, Lei e Poder Judicial – O Direito em Ação", em 2020, na Universidade Autonoma de Lisboa [pendente de publicação].

[280] *ARIZONA STATE UNIVERSITY – Soft-Law Governance of Artificial Intelligence, Executive Sumary*.

[281] MARCHANT, Gary - *"Soft Law" Governance of Artificial Intelligence*, p. 5.

[282] Eis as clássicas "Leis da Robótica" formuladas por ISAAC AZIMOV: "1ª Lei: Um robô não pode ferir um ser humano ou, por omissão, permitir que um ser humano sofra algum mal. 2ª Lei: Um robô deve obedecer as ordens que lhe sejam dadas por seres humanos, exceto nos casos em que tais ordens entrem em conflito com a Primeira Lei" (AZIMOV, Isaac - *I, Robot*).

[283] Não à toa diversos instrumentos de *soft law* fazem expressa referência às leis enunciadas pelo literato e cientista russo. Entre eles, pode ser mencionado o conjunto de Recomendações do Parlamento Europeu à Comissão sobre disposições de Direito Civil sobre Robótica, em cujos *considerandas* lê-se: "[...] U. Considerando que as Leis de Asimov têm de ser perspetivadas como sendo direcionadas aos criadores, produtores e operadores de robôs, incluindo robôs com autonomia integrada e autoaprendizagem, uma vez que aquelas leis não podem ser convertidas em código de máquina [...]". PARLAMENTO EUROPEU - Recomendações do à Comissão sobre disposições de Direito Civil sobre Robótica.

Mais recentemente, em uma das primeiras manifestações estruturadas e oficiais, o Governo da Coreia do Sul anunciou o processo de redação de uma "Carta sobre Ética Robótica", em 2007[284].

Destinada a prevenir "doenças sociais" possivelmente resultantes de "medidas legais inadequadas"[285], a Carta sul-coreana compõe-se de três partes: i) Padrões de desenvolvimento ("*Manufacturing Standards*") – entre os quais se destaca a imposição de controle de qualidade da produção, proteção de dados pessoais e restrições à autonomia dos robôs; ii) Direitos e deveres dos usuários e proprietários ("*Rights & Responsibilities of Users/Owners*") – com o asseguramento de direitos como o de controle dos robôs, privacidade e segurança, ao lado de deveres como o de abstenção de seu uso para atividades ilícitas ou que provoquem danos físicos ou psicológicos a outras pessoas; e iii) Direitos e deveres dos robôs ("*Rights & Responsibilities for Robots*") – claramente inspirados no texto de AZIMOV, os deveres referem-se às obrigações de não ferir, por ação ou omissão, seres humanos, de obedecer seus usuários ou proprietários (salvo se daí puder resultar ofensa ao primeiro dever), além de não enganá-los[286].

Destaque-se que, curiosamente, o documento prevê dois direitos aos robôs: o de existir, sem medo de "ferimentos" ou da "morte" e o de viver uma existência livre de abusos sistemáticos[287].

3.1.2.1. Iniciativa Global do IEEE sobre Ética de Sistemas Autonomos e Inteligentes

Em 2016, uma abrangente proposta de uso de *soft law* para regular o fenômeno da IA conduzida pelo Instituto de Engenharia Elétrica e Eletrônica (IEEE) resultou no lançamento da "Iniciativa Global do IEEE sobre Ética de Sistemas Autonomos e Inteligentes" ("*IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*") [288].

---

[284] MARCHANT, Gary – *Op. cit.*, p. 5.

[285] No original: "*This Charter was drafted in order to prevent social ills that may arise out of inadequate social and legal measures to deal with robots in society*". A íntegra de uma tentativa de reprodução não oficial do documento pode ser acessada em: https://docs.google.com/document/d/1aoF36DW2F-D-VWve6R4cMK-0BZ58vyHm-eH3F_LzTS4/edit [Consult. 8 jan. 2020]..

[286] Fonte: https://docs.google.com/document/d/1aoF36DW2F-D-VWve6R4cMK-0BZ58vyHm-eH3F_LzTS4/edit [Consult. 8 jan. 2020]..

[287] No original: "*Under Korean Law, Robots are afforded the following fundamental rights: i) The right to exist without fear of injury or death. ii) The right to live an existence free from systematic abuse*". Fonte: https://docs.google.com/document/d/1aoF36DW2F-D-VWve6R4cMK-0BZ58vyHm-eH3F_LzTS4/edit [Consult. 8 jan. 2020]..

[288] MARCHANT, Gary – *Op. cit.*, p. 5. O inteiro teor encontra-se disponível em: https://standards.ieee.org/industry-connections/ec/autonomous-systems.html [Consult. 8 jan. 2020]..

Endereçada, em última análise, aos desenvolvedores de sistemas de IA, a iniciativa busca assegurar-lhes formação ética com o fito de direcionar os avanços tecnológicos em prol da humanidade.

Houve dois principais frutos concretos da Iniciativa do IEEE: i) um guia conhecido como "Design Eticamente Alinhado" [289]; e ii) um conjunto de "standards" voltado para governança e aspectos éticos da IA[290].

O guia contém mais de uma centena de indicações relacionadas a aspectos legais e éticos, além de políticas concernentes ao desenvolvimento de sistemas de IA. Resulta de uma preocupação que vai além da busca pela mera eficiência técnica de tais sistemas, com ênfase no direcionamento finalístico para se estabelecerem propósitos benéficos ao homem e ao meio ambiente[291].

Foi resultado das contribuições de cerca de 250 especialistas voltadas ao alinhamento pragmático dos sistemas de IA a valores e princípios éticos[292], centrada em uma perspectiva eudemônica (no sentido aristotélico da expressão) no bem-estar humano em um "dado contexto cultural", com respeito às diferentes tradições morais[293].

Inspiraria, de outra parte, a criação de programas de certificação e instigaria a adoção de políticas nacionais e globais em linha com tais valores e princípios[294].

Seriam oito os "princípios" delineados no guia do IEEE: direitos humanos, bem-estar, proteção de dados, efetividade, transparência, responsabilidade, precaução e competência técnica[295].

Já os "stantards"[296] ainda se encontram em processo de elaboração, com conclusão prevista para 2021[297]. Concernem à governança e aspectos éticos de IA em um espectro amplo de segmentos que compreende mais de uma dezena de pontos, entre os quais podem ser destacados: transparência dos sistemas autônomos (IEEE P7001), privacidade de dados (IEEE P7002), enviesamento de dados (IEEE P7003), transparências na governança de dados laborais (IEEE P7005), direcionamento ontologicamente ético de robôs e sistemas

[289] IEEE – *Ethically Aligned Design*.
[290] A descrição encontra-se disponível em: https://standards.ieee.org/industry-connections/ec/autonomous-systems.html [Consult. 8 jan. 2020].
[291] IEEE – *Ethically Aligned Design*, p. 2.
[292] MARCHANT, Gary – *Op. cit.*, p. 5.
[293] IEEE – *Op. cit.*, p. 2. A construção de bases de consenso moral talvez represente uma das maiores dificuldades para o estabelecimento de diretrizes universais para regular a IA (ou qualquer outra tecnologia, em verdade), dada a heterogeneidade perceptível de tais tradições.
[294] MARCHANT, Gary – *Op. cit.*, p. 6.
[295] IEEE – *Op. cit.*, p. 4.
[296] Da observância de tais padrões resultaria a certificação (série IEEE P7000) das entidades supervio
[297] MARCHANT, Gary – *Op. cit.*, pp. 6-7.

autônomos (IEEE P7007), métricas de bem-estar (IEEE P7010) e processos de identificação e mensuração de confiabilidade (IEEE P7011)[298].

3.1.2.2. Princípios de Asilomar

A Conferência de Asilomar[299] sobre "Benefical AI" ocorreu em 2017, entre os dias 5 e 8 de janeiro, sob os auspícios do Instituto para o Futuro da Vida ("*Future of Life Institute – FLI*")[300].

Contou com especialistas de vários seguimentos e abordou temas como: a) "Economia - como aumentar nossa prosperidade através da automação sem deixar pessoas sem renda ou propósito?"; b) "Criando inteligência artificial em nível humano: isso acontecerá e, em caso afirmativo, quando e como? Que principais obstáculos remanescentes podem ser identificados? Como podemos tornar os futuros sistemas de inteligência artificial mais robustos do que os de hoje, para que eles façam o que quisermos sem ferir, funcionar mal ou ser invadido?"; c) "Superinteligência: Ciência ou ficção? Se a IA geral de nível humano é desenvolvida, então quais são os resultados prováveis? O que podemos fazer agora para maximizar a probabilidade de um resultado positivo?"; e d) "Lei, política e ética: Como podemos atualizar os sistemas legais, os tratados e os algoritmos internacionais para serem mais justos, éticos e eficientes, e para manter o ritmo com a IA?"[301].

O conjunto compreende 23 princípios, agrupados em três subconjuntos: i) questões de pesquisa; ii) ética e valores; e iii) questões de longo prazo.

Entre as questões relativas a pesquisas, podem ser destacados as seguintes diretrizes: a) a sua meta deve ser a criação de inteligência benéfica; b) seria necessário um intercâmbio saudável e construtivo entre pesquisadores e formuladores de políticas públicas; c) a cultura da cooperação, confiança e transparência deveria ser fomentada entre pesquisadores e desenvolvedores de sistemas de IA; e d) cumpriria evitar-se uma espécie de corrida, em que a competição entre os atores pudesse vir a comprometer os padrões de segurança.

---

[298] MARCHANT, Gary – *Op. cit.*, p. 7.

[299] Foi aí também que em 1975 se deu importante conferência sobre engenharia genética, a "*Conference on Recombinant DNA*", na qual se gestou um instrumento de *soft law* contendo diversas orientações gerais sobre o tema. Maiores informações sobre o evento podem ser encontradas em https://www.ncbi.nlm.nih.gov/books/NBK234217/ [Consult. 2 jan. 2020].

[300] Informações extraídas de https://futureoflife.org/2017/01/17/principled-ai-discussion-asilomar/ [Consult. 25 jun. 2019].

[301] Fonte: https://futureoflife.org/bai-2017/?cn-reloaded=1 [Consult. 25 jun. 2019].

No tocante aos temas "ética" e "valores" em que se concentram os aspectos mais relevantes do ponto de vista estritamente jurídico, sobressaem os princípios a seguir: a) segurança – durante todo o seu ciclo de vida operacional e mediante padrões verificáveis; b) transparência – para expor suas vulnerabilidades e, quando se tratar de ferramenta jurisdicional, para prover explicações auditáveis por autoridade humana competente; c) responsabilidade – de modo a permitir a imposição de sanções em face de implicações decorrentes de falhas ou mau uso; d) alinhamento de valores – as metas dos sistemas deveriam ser alinhadas aos valores humanos fundamentais, v.g., dignidade da pessoa humana, liberdade, diversidade cultural; e) privacidade – com direito de acesso a todos os dados pessoais eventualmente coletados e armazenados; f) benefícios e prosperidade compartilhados – compartilhamento universal dos bens gerados pelos sistemas de IA; g) controle humano – o nível de delegação aos agentes artificialmente inteligentes deveria ser objeto de decisão humana; e h) não subversão – o poder conferido aos sistemas de IA deve respeitar e incrementar, nunca subverter os processos cívicos e sociais.

Em relação às questões de longo prazo, em que subsiste uma densa névoa no concernente aos cenários futuros ainda predominantemente especulativos, podem ser realçados: a) cautela quanto à capacidade dos sistemas – enquanto inexistir algum consenso científico, dever-se-ia evitar presunções quanto às limitações dos sistemas de IA a serem desenvolvidos; b) importância – a relevância disruptiva dos sistemas de IA deve atrair preocupações quanto à sua gestão e planejamento cuidadosos; c) riscos – as potencialidade catastróficas ("apocalípticas") não podem ser menosprezadas; d) auto-aprimoramento recursivo – sistemas que contem com a capacidade de se auto-aprimorarem de maneira recursiva devem se sujeitar a controles de seguranças mais rigorosos; e e) bem comum – a superinteligência artificial deve ser colocada a serviço de ideais éticas compartilhadas universalmente, não podendo ser apropriadas por entidades públicas ou privadas.

Até aqui a declaração de princípios de Asilomar já teria sido subscrita por 1.583 pesquisadores da área e IA e robótica, além de outras 3.447 pessoas com perfis diversos[302].

Trata-se, como se vê, de um instrumento cuja influência precípua restringe-se quase que exclusivamente ao âmbito acadêmico, mas que tem sido amplamente difundido, graças sobretudo aos esforços empreendidos pelo FLI[303].

[302] Fonte: https://futureoflife.org/ai-principles/?cn-reloaded=1 [Consult. 2 jan. 2020].

[303] Não obstante, vale observar que a Declaração de Asilomar atuou como fonte material de *hard law* ao ser transposta formalmente para a legislação do Estado da Califórnia nos Estados Unidos. STATE OF CALIFORNIA - Assembly Concurrent Resolution No. 215 (Sept. 7, 2018). MARCHANT, Gary – *Op. cit.*, p. 14.

Diante da circunstância de que uma parte significativa dos sistemas ainda se encontra em fase embrionária e diversos avanços pendem de pesquisas científicas mais aprofundadas ou mesmo de linhas a serem inauguradas, a Declaração de Asilomar pode ser reputada relevante para induzir a comunidade científica a observar os parâmetros supramencionados.

3.1.2.2. Recomendação da OCDE sobre Inteligência Artificial

3.1.2.2.1. Origem

A Recomendação da OCDE sobre Inteligência Artificial foi aprovada em 22 de maio de 2019 pelo Conselho Ministerial, a partir de proposta formulada pelo Comitê de Política Econômica Digital (CPED)[304]. Pode ser considerada precursora, por ter sido o primeiro produto normativo intergovernamental cujo objeto central consistiu nessa temática específica[305].

O instrumento teve como fim último a conciliação entre inovação e confiabilidade no desenvolvimento de sistemas de IA, a partir da promoção do respeito aos direitos humanos e aos valores democráticos[306].

Atua em complemento aos "*standards*" já definidos em outras áreas adjacentes como é o caso da gestão de riscos concernentes à segurança digital, privacidade e conduta empresarial responsável[307].

A preocupação da OCDE foi a de estabelecer parâmetros suficientemente flexíveis de modo a facilitar a sua implementação e garantir-lhes alguma longevidade, considerando o cenário tecnológico especialmente cambiante no campo da IA[308].

Além de contar, naturalmente, com a adesão de todos os 36 países membros da OCDE (o que inclui, como se sabe, Portugal), a Recomendação obteve mais 8 aderentes por parte de não-membros, entre os quais figura o Brasil, que aderiu ao instrumento em 21/05/19[309].

---

[304] Informações extraídas de https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 [Consult. 8 jul. 2019].
[305] *Id. ibid.*
[306] *Id. ibid.*
[307] *Id. ibid.*
[308] *Id. ibid.*
[309] Além do Brasil (que aderiu antes mesmo da aprovação pelo Conselho Ministerial da OCDE), aderiram na mesma data Argentina, Colômbia, Costa Rica, Peru e Romania. Em 19/12/19, Malta aderiu e, em 29/10/19, a Ucrânia. Informações extraídas de https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 [Consult. 8 jul. 2019].

3.1.2.2.2. Atores

A Recomendação foi o resultado da atuação coordenada e multidisciplinar decorrente da contribuição de mais de 50 especialistas de diferentes segmentos, que se reuniram entre os meses de setembro de 2018 e fevereiro de 2019 em diversos países (França, Estados Unidos e nos Emirados Árabes) [310].

Entre outros havia representantes[311]: i) do governo – o que incluiu diversos ministros, entre os quais podem ser citados o Ministro da Ciência e Tecnologia da Coréia do Sul, o Ministro da Economia e do Trabalho da Finlândia e o Ministro da Economia e do Meio Ambiente da Holanda; ii) da indústria – quanto aos quais podem ser lembrados os diretores de empresas líderes do seguimento de tecnologia da informação e comunicação, e.g., Google, Facebook, IBM e Microsoft; iii) da sociedade civil – que se fez presente por meio de executivos e dirigentes de instituições como "*AI Iniciative*"[312], "*World Privacy Forum*"[313] e "*OpenAI*"[314]; iv) de entidades sindicais – as quais foram, por sua vez, representadas pela "*UNI Global Union*"[315]; assim como v) da comunidade acadêmica – com pesquisadores e estudiosos de prestigiadas instituições de ensino, tais como a Universidade de Harvard, o Instituto de Tecnologia de Massachusetts (MIT) e a "*University College London*" (UCL).

---

[310] *Id. ibid.*

[311] *Id. ibid*. A relação nominal de todos os colaboradores encontra-se em http://www.oecd.org/going-digital/ai/oecd-aigo-membership-list.pdf [Consult. 8 jul. 2019].

[312] Eis a descrição dos objetivos da entidade: "*The AI Initiative a été lancée en 2015 dans le cadre de « The Future Society » pour se concentrer sur les questions d'impact et de gouvernance de la montée en puissance de l'Intelligence Artificielle. Notre objectif est de favoriser l'émergence de cadres de gouvernance globaux permettant de tirer le meilleur parti de la révolution de l'IA en maximisant les bénéfices et en minimisant les risques*". Fonte: http://theconversation.com/the-ai-initiative-conversation-avec-nicolas-miailhe-89582 [Consult. 10 jul. 2019]. Informações mais detalhadas podem ser encontradas em http://thefuturesociety.org/the-ai-initiative/ [Consult. 10 jul. 2019].

[313] Segundo as descrições de seu sítio oficial, o "Fórum Mundial de Privacidade" dedicar-se-ia a "reimaginar a privacidade na era digital por meio de pesquisas, análises aprofundadas sobre o tema, além de oferecer ao consumidor educação da mais alta qualidade" [tradução livre]. Sua visão seria a de "capacitar as pessoas com o conhecimento sobre direitos e ferramentas necessárias para proteger sua privacidade e moldar sua vida digital a partir daí" [tradução livre]. Fonte: https://www.worldprivacyforum.org/about-us/ [Consult. 10 jul. 2019].

[314] Trata-se de uma instituição sem fins lucrativos de pesquisa em inteligência artificial, baseada em São Francisco e que tem como escopo a promoção e o desenvolvimento de sistemas de IA amigáveis, destinados a produzir benefícios a humanidade como um todo. Fonte: https://openai.com/about/ [Consult. 10 jul. 2019].

[315] Com sede na Suiça, a UNI representaria, por meio das entidades sindicais a ela vinculadas, cerca de 20 milhões de trabalhadores de mais de 150 países. Fonte: https://www.uniglobalunion.org/about-us-0 [Consult. 10 jul. 2019].

3.1.2.2.3. Conteúdo do instrumento

O instrumento pode ser decomposto para fins estritamente didáticos, nas seguintes partes fundamentais: i) considerações preambulares; ii) conceitos essenciais; iii) princípios estruturantes (Seção 1); e iv) políticas nacionais e cooperação internacional (Seção 2) [316].

Em suas considerações iniciais, procura-se situá-la no contexto da Agenda de 2030 para o Desenvolvimento Sustentável adotada pela Assembleia Geral Organização das Nações Unidas (ONU)[317], assim como da Declaração Universal dos Direitos Humanos de 1948[318].

Em seguida, menciona as implicações globais da IA em diversos setores, com ênfase na dimensão econômica e no mundo do trabalho, com potencial de ampliação da riqueza e do bem-estar a partir do incremento da inovação e da produtividade[319].

Cumpre destacar, nesse particular, o quadro assimétrico que se descortina no horizonte, diante da óbvia heterogeneidade nos estágios de desenvolvimento nos quais se encontram os países, a ensejar uma natural diversidade nos impactos sentidos por tais nações quanto às tecnologias relativas à IA[320].

Reporta-se, ainda em suas notas preambulares, à necessidade de se garantir confiança como premissa para que as transformações digitais possam se difundir de maneira mais abrangente, ao assegurar a exploração das potencialidades benéficas e conter os riscos emergentes da disrupção tecnológica em curso[321].

---

[316] Informações extraídas de https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 [Consult. 8 jul. 2019].

[317] De acordo com o texto oficial, desenvolvimento sustentável seria aquele capaz de "atender às necessidades do presente sem comprometer a capacidade das gerações futuras de atender às suas próprias necessidades" (https://www.un.org/sustainabledevelopment/development-agenda/ - Consult. 15 nov. 2019). Trata-se, como se vê, de conceito bastante pertinente, considerados, de um lado, as potencialidades benéficas da IA para toda a humanidade, e os riscos de comprometimento de valores humanos centrais, de outro.

[318] A relação entre direitos humanos e IA revela possibilidades de simbiose promissoras, mas abriga, outrossim, tensões importantes com virtuais entrechoques. Ilustra o primeiro aspecto a utilização de ferramentas artificialmente inteligentes com o objetivo de endereçar problemas complexos e de alcance global, v.g., analfabetismo, fome, crise energética, desequilíbrios ecológicos – a esse propósito, para uma noção de projetos concretos em tais segmentos vide https://ai4good.org/active-projects/ [Consult. 12 ago. 2019]. Já o segundo refere-se tanto aos riscos decorrentes dos futuros desenvolvimentos tecnológicos (como é o caso das já mencionadas armas letais autônomas) como das soluções já em uso – v.g., sistemas que se apropriam de dados massivos para dar suporte a decisões em vários segmentos, por vezes a comprometer direitos fundamentais, tais como, isonomia e privacidade, para citar apenas dois exemplos mais evidentes.

[319] Informações extraídas de https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 [Consult. 8 jul. 2019].

[320] Aqui pode vir a se dar o fenômeno do "*winner takes all*", em que os estados (ou empresas) a conquistarem de forma pioneira a desenvoltura na criação e manuseio de aludidas ferramentas podem adquirir tal vantagem competitiva a ponto de inviabilizarem a concorrência de outros, de modo a eventualmente subjulgá-los tecnológica e economicamente. Entre outras fontes: https://becominghuman.ai/in-the-battle-for-artificial-intelligence-winner-takes-all-4c2447e68237 [Consult. 18 jul. 2019].

[321] Informações extraídas de https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 [Consult. 8 jul. 2019].

Ressalta, de outro lado, a relevância do estabelecimento de marcos normativos a partir de uma abordagem condizente com a natureza intrinsecamente volátil do fenômeno, de forma a tutelar os direitos humanos, a propriedade intelectual, a privacidade, a competitividade, entre outros[322].

Reconhece, igualmente, a conveniência de se desenhar um ambiente com políticas estáveis que promovam uma abordagem centrada no ser humano para estimular pesquisas e incentivos econômicos à inovação em linha com tal pressuposto, a vincular todos os "*stakeholders*" envolvidos[323].

Enaltece-se, ademais, a confiabilidade como parâmetro competitivo global dirigido a todos os atores que venham a contribuir para enfrentar os desafios resultantes das aplicações de IA[324].

A Recomendação descreve, de outro lado, cinco conceitos fundamentais: a) sistema de IA ("*AI system*"); b) ciclo de vida do sistema de IA ("*AI system lifecycle*"); c) conhecimento da IA ("*AI knowledge*"); d) atores da IA ("*AI actors*"); e e) partes interessadas ("*Stakeholders*") [325].

Um "sistema de IA" encontra-se descrito como um sistema computacional que pode, para um determinado conjunto de objetivos definidos pelo homem, fazer previsões, recomendações ou mesmo tomar decisões que influenciam ambientes reais ou virtuais, com diferentes níveis de autonomia.

Já o "ciclo de vida do sistema de IA" seria composto das seguintes fases, não necessariamente sequenciais: i) "design" – abrangeria o planejamento, coleta e processamento de dados, assim como a construção do modelo; ii) verificação e validação; iii) implantação; e iv) operação e monitoramento.

O "conhecimento de IA" refere-se às habilidades e recursos, como dados, códigos-fontes, algoritmos, modelos, pesquisas, programas de treinamento, governança, processos e melhores práticas, necessários à compreensão e intervenção em qualquer das fases do ciclo de vida do sistema de AI.

"Atores de IA" consistiriam naqueles que desempenham papel ativo no seu ciclo de vida, o que compreende organizações e indivíduos responsáveis pela implementação ou operação do sistema.

---

[322] *Id., ibid*.
[323] *Id., ibid*.
[324] *Id., ibid*.
[325] *Id., ibid*.

"Partes interessadas" seriam todas as organizações e indivíduos envolvidos, ou afetados por sistemas de IA, direta ou indiretamente, de modo que os "atores de IA" poderiam ser tidos como um subconjunto seu.

Cinco núcleos principiológicos fundantes e lastreados em valores são identificados pela Recomendação, que conclama todos os atores de IA a promovê-los e implementá-los para ensejar uma gestão responsável da confiabilidade dos sistemas de IA: a) crescimento inclusivo, desenvolvimento sustentável e bem-estar ("*inclusive growth, sustainable development and well-being*"); b) valores centrados no ser humano e justiça ("*human-centred values and fairness*"); c) transparência e explicabilidade ("*transparency and explainability*"; d) robustez, proteção contra falhas e segurança ("*robustness, security and safety*"); e e) responsabilidade ("*accountability*")[326].

De acordo com o primeiro núcleo ("crescimento inclusivo, desenvolvimento sustentável e bem-estar"), as partes interessadas deveriam se engajar em uma gestão proativa dos sistemas de IA na busca de resultados positivos para as pessoas e para o meio ambiente[327]. Entre os exemplos citados na própria Recomendação podem ser mencionados: o aumento das capacidades humanas, a inclusão de populações subrepresentadas e a redução das desigualdades econômicas, sociais e relativas a gênero[328].

No tocante à centralidade do ser humano e à justiça, os atores de IA deveriam respeitar a Lei, os direitos humanos (v.g., liberdade, dignidade, autonômica, privacidade, igualdade, garantias laborais) e os valores democráticos em todas as fases do ciclo de vida dos sistemas[329].

Para respeitar tais diretrizes, implementariam [os atores de IA] mecanismos de proteção que permitissem resguardar os referidos direitos e valores, de acordo com o contexto e o estado da arte dos sistemas de IA[330].

Os atores de IA se comprometeriam, ainda, a garantir transparência dos sistemas[331], providenciando informações inteligíveis quanto aos seguintes aspectos: a) seu modo de

---

[326] *Id., ibid*.
[327] *Id., ibid*.
[328] *Id., ibid*.
[329] *Id., ibid*.
[330] *Id., ibid*.
[331] A opacidade dos sistemas de IA apresenta-se como uma das delicadas e complexas características intrínsecas à maior parte dos modelos. BAROCAS e SELBST assim descrevem-na: "*This inscrutability is a natural fact of machine learning systems derived from their inherent complexity. As Jenna Burrell has noted: Though a machine learning algorithm can be implemented simply in such a way that its logic is almost fully comprehensible, in practice, such an instance is unlikely to be particularly useful. Machine learning models that prove useful (specifically, in terms of the 'accuracy' of classification) possess a degree of unavoidable complexity*". BAROCAS, Solon and SELBST, Andrew D. - *Regulating Inscrutable Systems*, pp. 13-14.

funcionamento; b) existência de interações com agentes artificialmente inteligentes inclusive no local de trabalho; e c) bases de dados utilizada e lógica subjacente às predições, análises, recomendações e decisões aptos a permitirem aos negativamente afetados a impugnação dos resultados[332].

À luz do quarto corpo de princípios, seria exigido dos sistemas de IA robustez, proteção contra falhas e segurança, de modo a evitar riscos decorrentes de mau funcionamento[333].

Para atingir semelhante desiderato, além de manter um programa contínuo de gestão de riscos compatível com seu papel, os atores de IA assegurariam rastreabilidade nas bases de dados, processo e decisões em todas as etapas do ciclo de vida dos sistemas, para permitir uma análise dos resultados decorrentes de sua execução e obstar falhas que comprometam a privacidade, segurança digital ou a imparcialidade, de forma a impedir o "enviesamento algorítmico" ("*algorithm bias*")[334].

Por fim, de forma condizente com a função desenvolvida, todos os atores deveriam ser passíveis de imputação de responsabilidade pelos danos ocasionados por eventual inobservância dos supramencionados princípios[335].

3.1.2.2.4. Políticas nacionais e cooperação internacional

A OCDE recomenda aos Estados aderentes, com ênfase em pequenas e médias empresas, a realização de investimento público, assim como o estímulo a investimento privado, em pesquisas interdisciplinares para alavancar a confiabilidade na IA, a partir do enfrentamento de questões técnicas, legais e éticas[336].

Exorta, ainda, a conceder preferência a bases de dados abertas, que sejam representativas, respeitem à privacidade e estejam livres do "enviezamento algorítmico", além de incrementar a interoperabilidade entre os diversos sistemas de IA[337].

---

[332] *Id., ibid*.
[333] *Id., ibid*.
[334] *Id., ibid*. O tema vem sendo explorado por diversos autores, entre os quais pode ser citado PENEL, que apresenta a questão do seguinte modo: "*The issue is that bias is not only transferred to algorithms, but it can also be amplified. This happens when biased ML algorithms create new data that is re-injected into the model as part of its continuous training. It becomes worse when the biased algorithm is used to make millions of predictions per minute as part of an automatic decision-making process, bringing bias back into the real world, at scale*". PENEL, Olivier - *Algorithms, the Illusion of Neutrality: The Road to Trusted AI*, p. 3.
[335] *Id., ibid*.
[336] *Id., ibid*.
[337] *Id., ibid*.

Indica a necessidade de os Governos desenvolverem um "ecossistema digital" que abrigue de forma confiável tais sistemas, mediante a implementação de mecanismos para garantir o compartilhamento seguro, justo, legal e ético de dados[338].

Instiga, de outro lado, a adoção de testes controlados para submeter os sistemas de IA a experimentos que avaliem sua escalabilidade[339].

Além disso, também sugere a avaliação da necessidade de ajustamento das arquiteturas regulatórias e das políticas públicas com o objetivo de encorajar a inovação e competitividade no desenvolvimento de sistemas confiáveis de IA[340].

Quanto às transformações no mercado de trabalho, as iniciativas governamentais deveriam considerar a habilitação ou requalificação dos trabalhadores para manusear as novas tecnologias relativas à IA, o diálogo social para assegurar transição justa, ante a supressão de postos de trabalho e a criação de novas oportunidades profissionais. Nesse mesmo sentido também orienta a promoção do uso responsável da IA no mundo do trabalho, para ampliar a segurança laboral e a qualidade dos empregos, de um lado, e promover o empreendedorismo e produtividade, de outro, sempre de modo a ensejar o compartilhamento abrangente e justo dos benefícios resultantes da IA[341].

Por fim, aponta a cooperação internacional mesmo entre países em desenvolvimento como instrumento para fazer cumprir os princípios e implementar as políticas supramencionados na gestão da confiabilidade da IA. O trabalho conjunto no âmbito na OCDE e em outros fóruns internacionais deveria prestigiar o diálogo transetorial e aberto para angariar "expertise" compartilhada nesse segmento, bem como estimular a adoção de padrões de interoperabilidade globais frutos de consenso. Seria, ainda, relevante a criação de uma métrica internacional para avaliar o progresso na implementação dos princípios delineados pela Recomendação na pesquisa, desenvolvimento e implementação de sistemas de IA[342].

[338] *Id., ibid*.
[339] *Id., ibid*.
[340] *Id., ibid*.
[341] *Id., ibid.* Entre os avanços identificados nessa direção, pode ser mencionado o conjunto de "standards" que se encontra em fase de desenvolvimento por parte do Instituto de Engenharia Elétrica e Eletrônica (IEEE), conforme melhor se examinará adiante.
[342] *Id., ibid*.

3.1.2.2.5. "Ideia-força"

Confiabilidade e abordagem centrada no ser humano talvez possam ser reputadas as principais "ideias-forças" da Resolução.

Efetivamente, a ideia de confiança ou confiabilidade permeia diversas das linhas principiológicas da Recomendação e subjaz como lastro axiológico primário de princípios como o da transparência, robustez e responsabilidade.

Além de mirar na construção de uma plataforma confiável sob as perspectivas técnica, ética e jurídica, a OCDE também parece considerar a confiabilidade sob a perspectiva econômica, sobretudo ante as fontes de instabilidade resultantes do desenvolvimento e implementação de sistemas de IA nas relações entre agentes econômicas.

Já a centralidade do ser humano figura como matriz fundante dos princípios relativos ao crescimento inclusivo, desenvolvimento sustentável e bem-estar, assim com ao respeito aos direitos humanos e aos valores democráticos.

Com isso, busca-se restituir a fundamentalidade de tais direitos e valores e afirmar o caráter instrumental das ferramentas tecnológicas.

3.1.3. Autoregulação[343]

As formas de regulação tradicionais, sobretudo na perspectiva europeia, envolvem a ideia de "controlo por um superior", de forma a cumprir uma "função diretiva"[344].

Portanto, como se sabe, representa uma excecionalidade a possibilidade de os entes estatais se absterem de regular determinados fenómenos com relevância social e "tolerarem" ou mesmo estimularem a regulação promovida pelos próprios agentes interessados.

Isso se deu em campos nos quais a complexidade inerente levou ao encorajamento de tentativas autoregulatórias, v.g. quanto à radio e teledifusão, assim como ao mercado publicitário[345], além da indústria nuclear e do segmento de nanotecnologia, dentre outros[346].

---

[343] Algumas das considerações contidas nesse item ("3.1.3. Autoregulação") foram extraídos de: PAULA, Gáudio Ribeiro de – A Recomendação da OCDE Sobre Inteligência Artificial e sua Relevância Como Instrumento de "Soft Law" para Contenção dos Riscos aos Direitos Fundamentais. Artigo apresentado como exigência parcial para obtenção do Grau de Doutoramento em Direito na disciplina "Estado, Lei e Poder Judicial – O Direito em Ação", em 2020, na Universidade Autonoma de Lisboa [pendente de publicação].

[344] OGUS, A. *Apud* KLEINSTEUBER, Hans J. - *Self-regulation, Co-regulation, State Regulation – The Internet between Regulation and Governance*, p. 63.

[345] KLEINSTEUBER, Hans J. - *Self-regulation, Co-regulation, State Regulation – The Internet between Regulation and Governance*, p. 63.

[346] Fonte: https://www.regulation.org.uk/specifics-self-regulation.html [Consult. 23 set. 2022].

Atua, pois, em áreas com deficit regulatório, ao se apresentar como alternativa à ausência de regras estatais[347].

Há quem defenda o seu emprego na regulação da IA, ante a sua agilidade e capacidade de se ajustar às particularidades e complexidades ínsitas à tecnologia, de forma a compaginar segurança jurídica e inovação[348].

Para assegurar a efetividade de práticas autoregulatórias, seria necessário observar algumas características, entre as quais se poderia destacar: i) a formulação de conjuntos claros de princípios alinhados com os objetivos da industria (da IA); ii) os seus pesquisadores e desenvolvedores deveriam agir em função de motivações intrínsecas, para alcançar soluções relativas a ações coletivas; iii) existência de um "corpo" governativo dotado de autoridade e independência; iv) transparência interna e externa; e v) presença de mecanismos coertivos mínimos[349].

Entre as iniciativas de autoregulação promovidas por entes privados podem ser destacadas aquelas de grandes "*players*" do segmento da tecnologia da informação: Google, Microsoft e IBM.

Quanto ao Google, após destacar a responsabilidade decorrente de sua condição de líder em sua área de atuação, seu "*Chief Executive Officer*" (CEO) anunciou um conjunto sete princípios apresentados como objetivos ou metas a serem perseguidas no desenvolvimento de aplicações de IA. Ei-los: i) ser socialmente benéfico ("*be socially beneficial*"): ii) evitar criar ou reforçar enviezamentos injustos ("*avoid creating or reinforcing unfair bias*"); iii) ser construído e testado para garantir sua segurança ("*be built and tested for safety*"); iv) assegurar e prestação de contas, de modo a ser passível de atribuição de responsabilidade pessoal ("*be accountable to people*"); v) incorporar privacidade em seu design ("*incorporate privacy design principles*"); vi) manter padrões elevados de excelência científica ("*uphold high standards of scientific excellence*"); e vii) ser acessível apenas para uso de acordo com os princípios retromencionados ("*be made available for uses that accord with these principles*")[350].

---

[347] "*Self-regulation is an alternative form of governance where an industry designs and enforces new rules and standards for themselves, often in 'areas where government rules are lacking*' (Haufler 2001, 8)". LEENDERS, Gijs – *The Regulation Of Artificial Intelligence – A Case Study of the Partnership on AI*, p. 6.

[348] "*I posit that a self-regulatory system has the potential to be an effective and expedient solution to the unique regulatory challenges of AI, and could enable the AI industry to overcome the competitive dynamic that incentivises AI development speed over safety*". LEENDERS, Gijs – *Op. cit*, p. 6.

[349] LEENDERS, Gijs – *Op. cit*, p. 12.

[350] PICHAI, Sundar – *AI at Google: our principles*.

De outro lado, a companhia elencou quadro vedações, a atuarem como limitadores de seus projetos: i) sempre que houver risco de dano, serão incorporadas travas de segurança; ii) não será admitida a produção de armas ou outras ferramentas que produzam ofensas físicas diretamente ou permitam ferir pessoas; iii) são proibidas tecnologias que possibilitem a coleta de informações para vigilância em desobediência a normas internacionalmente aceitas; e iv) sistemas que desrespeitem princípios de direito internacional e direitos humanos seriam banidos[351].

Já a Microsoft relacionou, de modo bastante objetivo, seis princípios fundamentais. Todos decorreriam da preocupação de projetar sistemas de IA confiáveis, o que exigiria a criação de soluções que reflitam princípios éticos profundamente enraizados em valores importantes e atemporais.

Seriam eles: i) equidade ou justiça ("*fairness*") – devem tratar todas as pessoas de maneira justa; ii) inclusão ("*inclusiveness*") – devem capacitar todos e envolver as pessoas; iii) confiabilidade e segurança ("*reliability & safety*") – devem ter um desempenho confiável e seguro; iv) transparência ("*transparency*") – devem ser compreensíveis; v) privacidade e segurança ("*privacy & security*") – devem ser seguros e respeitar a privacidade ; vi) prestação de contas ("*accountability*") – devem ter responsabilidade algorítmica[352].

No caso da IBM, seu Presidente e CEO lembra que toda organização que desenvolve ou utiliza sistema de IA deve fazê-lo de forma responsável e transparente[353]. Ressalta que as empresas são julgadas sobretudo pelo modo como colhem e administram os dados coletados de outras pessoas[354].

Os três princípios encampados, nesse contexto, seriam: i) o propósito da IA deve ser o de ampliar a inteligência humana ("*the purpose of ai is to augment human intelligence*"); ii) dados e informações pertencem aos seus criadores ("*data and insights belong to their creator*"); e iii) as novas tecnologias devem ser transparentes e explicáveis ("*new technology, including ai systems, must be transparent and explainable*").

---

[351] PICHAI, Sundar – *Op. cit*.
[352] MICROSOFT – *Microsoft AI principles*.
[353] GINNI ROMETTY entende que a confiança seria um parâmetro essencial de julgamento das empresas por parte da sociedade. IBM – *IBM's Principles for Trust and Transparency*
[354] IBM – *IBM's Principles for Trust and Transparency*.

3.1.4. Co-regulação e Autoregulação regulada

A co-regulação pode ser definida como o resultado de um processo de diálogo entre as partes interessadas, que figura em um extrato intermediário entre: i) a "regulação de comando e controle estatal", diretamente, por meio de sua "central burocrática" ou, indiretamente, no exercício de "funções especializadas" (agências reguladoras independentes); e ii) a auto-regulação "pura", decorrente da atuação exclusiva dos entes regulados, tal como, de início, se observou na regulamentação dos padrões de infraestrutura da Internet[355].

Portanto, se o Estado e os reguladores privados cooperam em instituições conjuntas, pode-se considerar um quadro de "co-regulação"[356].

Se, contudo, esse tipo de autorregulação é estruturado pelo Estado, mas sem o seu envolvimento direto, o termo apropriado seria "autorregulação regulada". Este tipo de regulamento foi desenvolvido pela primeira vez na Austrália[357].

Ainda não se identificam, contudo, muitos exemplos relevantes dessa abordagem regulatória.

Uma das referências a serem lembradas seria a do Projeto de Lei brasileiro a ser adiante comentado (Projeto de Lei 21/2020), que poderia ser visto, em certo sentido, como um instrumento de autoregulação regulada.

Com efeito, o aludido projeto define uma estrutura principiológica e um conjunto de regras gerais e remete às instâncias autorregulatórias, explicitamente, a definição das normas específicas a disciplinar a questão.

[355] MARSDEN Christopher T. - *Co- and Self-regulation in European Media and Internet Sectors: The Results of Oxford University's Study*, p. 80.
[356] KLEINSTEUBER, Hans J. - *Op. cit.*, p. 63.
[357] KLEINSTEUBER, Hans J. – *Op. cit.*, p. 63.

3.1.5. Tecnoregulação[358]

3.1.5.1. Definições

BAYAMLIOĞLU e LEENES asseveram que o neologismo "tecno-regulação" (ou "*techno-regulation*") traduziria a "influência intencional do comportamento dos indivíduos ao incorporar normas nos sistemas e dispositivos tecnológicos"[359]. Cuida-se, assim, da regulação do comportamento humano a partir do uso de aparatos ("artefatos") tecnológicos nos quais se encontram inscritas regras de conduta e com os quais se poderia até mesmo inviabilizar (ou dificultar sobremaneira) fisicamente o descumprimento de tais normas.

Conceito próximo seria o de "captologia" ("*captology*"), que poderia ser considerado uma variação da tecno-regulação. Cuida-se de uma espécie de "ciência" cuja ênfase recairia sobre o "*design*", pesquisa, e análise de produtos computacionais interativos desenvolvidos com o objetivo mudar atitudes comportamentais. Estuda, pois, todos os modos de influenciar intencionalmente a conduta humana por meio de tais instrumentos[360].

A depender do contexto, os modelos regulatórios desse jaez podem ser referidos a partir das locuções "regulação pela tecnologia" ("*regulation by technology*"), "normatividade tecnológica" ("t*echnological normativity*"), "programas reguladores" ("*regulative software*"), "Direito como design" ("*law as design*"), "regulação baseada em design" ("*design-based regulation*") ou "regulação algorítimica" ("*algorithmic regulation*")[361].

O objeto da tecno-regulação pode compreender produtos, serviços, lugares ou pessoas, abarcando um espectro abrangente de práticas e designs[362]. Atualmente, podem ser citados como exemplos: o limitador de tempo de uso dos mecanismos de direção semiautônoma em carros mais modernos, os filtros de conteúdo na internet e os sistemas de controle de direitos autorais.

Não se trata, propriamente, de um fenômeno contemporâneo. Pode-se afirmar que ilustrariam historicamente essa abordagem tecno-regulatória: as correntes que prendiam os

---

[358] Boa parte das considerações aqui lançadas (item "3.1.5. Tecnoloregulação") foi extraído de relatório de autoria do subscritor da presente dissertação sob o título "Tecno-regulação, Inteligência Artifical e Estado De Direito Democrático" apresentado em 2018 como exigência parcial para obtenção do Grau de Doutoramento em Direito na disciplina "Direito: da Norma ao procedimento e à fase aplicativa", na Universidade Autonoma de Lisboa.
[359] No original: "*Techno-regulation refers to the intentional influencing of individuals' behaviour by embedding norms into technological systems and devices*". BAYAMLIOĞLU, Emre & LEENES, Ronald - *The 'rule of law' implications of data-driven decision- making: a techno-regulatory perspective*, p. 298.
[360] LEENES, Ronald E - *Op. cit.*, p. 154.
[361] BAYAMLIOĞLU, Emre & LEENES, Ronald - *The 'rule of law' implications of data-driven decision- making: a techno-regulatory perspective*, p. 298.
[362] BAYAMLIOĞLU, Emre & LEENES, Ronald - *Op. cit.*, p. 298.

escravos, constrangendo a sua liberdade de locomoção; os portões das cidades muradas para impedir o ingresso dos "inimigos" ou de qualquer pessoa a partir de determinado horário; as prisões destinadas a assegurar o cumprimento de penas restritivas de liberdade; e, mais recentemente, os "quebra-molas" ou limitadores eletrônicos de velocidade automotiva[363].

Seriam todos instrumentos de "regulação baseada em design" ("*design-based regulation*"), mecanismos de controle extremamente eficientes por agirem "*ex ante*" (obstando a conduta) e não "*ex post*" (sancionando a conduta), além de serem auto-executáveis por prescindirem de qualquer intervenção, conforme lembram MURRAY e SCOTT[364].

A tecnologia, nesse contexto, confundir-se-ia com a própria legislação, em certo sentido. É o que constata WINNER ao afirmar que "*technology in a true sense is legislation*". Assim, as formas tecnológicas moldariam, em grande medida, o padrão de conduta humano no nosso tempo, prossegue o autor[365].

Haveria, duas fontes tecno-regulativas (ou tecno-regulatórias), de acordo com BLACK: a) reguladores estatais - a tecnologia seria instrumento de implementação de normas provenientes do Estado (leis, em regra); e b) reguladores não estatais - agentes sem competência legiferante fariam uso de ferramentas tecnológicas para impor determinados comportamentos (poderiam ter força jurídica se houvesse respaldo contratual)[366].

3.1.5.2. Tecno-regulação na Internet

O advento da Internet e seus consectários despertou reflexões em torno dos efeitos jurídicos da arquitetura por meio da qual a rede mundial de computadores veio a estruturar-se. A disrupção catalisada pelas emergentes tecnologias de informação e comunicação trouxe

[363] RONALD LEENES, aludindo às contribuições de LANGDON WINNER, ilustra de modo curioso os efeitos normativos da arquitetura. De acordo com WINNER, o arquiteto de algumas pontes construídas em Long Island (Nova York) teria, propositalmente, limitado a altura dessas pontes a sete pés para impedir que os ônibus, predominantemente usados pela população negra relativamente pobre, alcançassem as praias. Daí conclui LEENES que "*Bridge-building thus is social engineering, because power relations are literally built into and perpetuated through stone*". LEENES, Ronald E - *Framing Techno-Regulation: An Exploration of State and Non-State Regulation by Technology*, p. 144.
[364] MURRAY, Andrew & SCOTT, Colin - *Controlling the New Media: Hybrid Responses to New Forms of Power*, p. 504.
[365] *Apud* LEENES, Ronald E - *Op. cit.*, p. 144.
[366] *Apud* LEENES, Ronald E - *Op. cit.*, p. 155.

ponderações acerca da própria natureza das normas derivadas ou inscritas no seu modelo de concepção[367].

Embora hoje tenham se tornado evidentes as implicações jurídicas de algumas das opções estruturantes adotadas para a conformação da "*World Wide Web*", inicialmente a comunidade jurídica parecia não se dar conta dos inúmeros riscos supervenientes aos direitos fundamentais[368].

Efetivamente, os únicos vetores de força que atuaram em um primeiro momento foram o Mercado (em primeiro lugar) e o Estado (em segundo). Releva destacar que as decisões fundamentais para a estruturação da Internet levaram em conta sobretudo critérios de ordem técnica e econômica, tendo sido olvidados os aspectos jurídicos, de início[369].

Entretanto, determinadas características da rede mundial de computadores atraíram a necessidade de uma análise jurídica do fenômeno sob o enfoque tecno-regulatório. Entre outras, podem ser lembradas: o relativo anonimato (a ensejar agressões eventuais a direitos como a honra), a ampla liberdade de circulação de bens imateriais (que permitiria possíveis lesões a direitos intelectuais) e a celeridade no tráfego de informações (com potencial lesivo à privacidade e mesmo à estabilidade do processo eleitoral - v.g. "fake news").

3.1.5.3. Aspectos jusfilosóficos

3.1.5.3.1. Planos dos "Ser" e "Dever ser" nas operações tecno-regulativas

Uma das objeções mais relevantes apresentadas por KELSEN ao jusnaturalismo concerne ao que o jusfilósofo tchéco-austríaco de "erro lógico fundamental"[370].

As características normativas da tecno-regulação parecem desafiar, em certo sentido, a clássica dicotomia adotada pelo autor da Teoria Pura do Direito ("*Reine Rechtslehre*") ao antepor "ser" (natureza) e "dever-ser" (sociedade).

---

[367] LESSIG afirma que as normas arquiteturais da Internet ostentariam uma natureza mais próxima a das leis físicas (mundo do "ser") que a das leis jurídicas (mundo do "dever-ser"). Isso porque sua inobervância seria quase tão inviável quanto a das leis naturais. LESSIG, Lawrence - *Code and other Laws of Cyberspace*.

[368] LESSIG, Lawrence - *Op, cit*.

[369] LESSIG, Lawrence - *Op, cit*.

[370] Eis a síntese da crítica delineada por ALEX SANDER PIRES: "A primeira objeção vem da definição de natureza, traduzida por Kelsen como ´a realidade empírica do acontecer fático em geral ou a natureza particular do homem tal qual ela se revela na sua conduta efetiva´. Em suma, o que se retira da natureza é um sucessão de fatos interligados que, pelo princípio da causalidade (causa-efeito), conduz a um resultado expectado - um ser -: ocorre que de um ´ser´ não se pode deduzir um ´dever-ser´, isto é, ´de um fato não se pode concluir uma norma, senão contrapô-los para que seja possível um juízo de valor". PIRES, Alex Sander Xavier - Justiça na perspectiva Kelseniana, p. 57.

Isso porque, nos artefatos tecnológicos construídos artificialmente pelo homem seriam "embebidas" normas de conduta cuja dinâmica aproximar-se-ia mais do plano do "ser" que o do "dever-ser".

Efetivamente, nessa perspectiva, do "ser" (artefato) brotariam preceitos impositivos ("dever-ser") ordenadores de conduta. Esse é o ponto de autores como o constitucionalista norte-americano e um dos teóricos precursores do debate jusfilosófico em torno da rede mundial de computadores, LAWRENCE LESSIG. Ao tratar da regulabidade da Internet, eis como explica, didaticamente, a regulação por meio do "código"[371]: "*The issue here is how the architecture of the Net—or its "code"—itself becomes a regulator. In this context, the rule applied to an individual does not find its force from the threat of consequences enforced by the law—fines, jail, or even shame. Instead, the rule is applied to an individual through a kind of physics. A locked door is not a command ´do not enter´ backed up with the threat of punishment by the state. A locked door is a physical constraint on the liberty of someone to enter some space"*[372].

De outro lado, tais elementos normativos inscritos em dispositivos poderiam inviabilizar (até mesmo materialmente) a inobservância do comando prescritivo neles "imbutido". Diante dessa possibilidade, aludidas normas estariam mais próximas das leis físicas ("ser") que das leis jurídicas ou morais ("dever-ser").

Entretanto, poder-se-ia objetar que, em verdade, aludidos artefatos seriam meros instrumentos de controle e execução de regras pre-definidas, não se confundindo com as normas propriamente ditas.

3.1.5.3.2. Livre arbítrio e autonomia

De acordo com MIREILLE HILDEBRANDT, haveria de duas modalidades de instrumentos tecno-regulatórios: a) os regulativos ("*regulative*") - que permitiriam a desobediência, criando apenas alguma espécie de dificultador da "não conformação" do comportamento (v.g. "quebra-mola" ou radar de velocidade); e b) os constitutivos

[371] O autor faz um trocadilho ao utilizar a expressão "*code*" em remissão, ao mesmo tempo, a código-fonte de programas de computador e a códigos enquanto produtos legislativos. LESSIG, LAWRENCE - *Code Version 2.0*, pp. 81-82. New York: Basic Books, 2006.
[372] LESSIG, LAWRENCE - *Op. cit.*, pp. 81-82.

("*constitutive*") - os quais obstariam de forma quase absoluta (por meio de constrição física, em regra) a adoção de conduta diversa da imposta (e.g. barreira de contenção na estrada) [373].

Nesse segundo caso, emerge a questão sobre o valor moral das ações ou omissões decorrentes de constrições impositivas. Seriam elas desprovidas de valor moral?

A esse propósito, revelam-se pertinente as lições de KANT recordadas por ALEX SANDER PIRES quanto à circunstância de a "liberdade ser condição da lei moral":

> Para que não se imagine encontrar aqui inconsequências, quando agora denomino a liberdade condição da lei moral e depois, no tratado, afirmo que a lei moral seja a condição sob a qual primeiramente podemos tornar-nos conscientes da liberdade, quero apenas lembrar que a liberdade é sem dúvida *ratio essendi* da lei moral, mas que a lei moral é a *ratio cognoscendi* da liberdade. Pois, se a lei moral não fosse pensada antes claramente em nossa razão, jamais nos consideraríamos autorizados a admitir algo como a liberdade (ainda que esta não se contradiga). Mas, se não existisse liberdade alguma, a lei moral não seria de modo algum encontrável em nós[374].

O que está em causa aqui é, em última análise, o livre arbítrio ou a autonomia, conceitos de alta relevância para a Ética, mas também para o Direito, sobretudo para os direitos humanos.

Consoante a leitura feita por BROWNSWORD[375] à luz do princípio da dignidade da pessoa humana, a autonomia deveria ser levada às últimas consequências, de forma a ensejar a escolha livre e consciente de todo sujeito de direitos em cumprir ou não as determinações legais emanadas do Estado. Ninguém poderia, pois, ser compelido a observar as normas estatais coercitivamente, sem contar com qualquer possibilidade de descumprimento.

Ainda de acordo com BROWNSWORD, a autonomia resultante da dignidade da pessoa humana implicaria três desdobramentos: a) o reconhecimento da capacidade de fazer as próprias escolhas; b) o respeito às escolhas que alguém faz livremente; e c) a necessidade de um contexto de apoio para a tomada de decisão (e ação) autônoma[376].

Todas essas dimensões estariam, a rigor, comprometidas diante da utilização de instrumentos tecno-regulatórios constitutivos, segundo o autor[377].

---

[373] *Apud* LEENES, Ronald E - *Op. cit.*, p. 150. Sobreleva notar a existência de autores que apenas admitem como tecno-regulação essa segunda espécie. É o caso de BROWNSWORD para quem a sua definição seria: "*a design-modality which prevents harm-generating behaviour in its entirety by overriding human decision and action*". *Apud id., ibid*.

[374] KANT, Immanuel. Apud PIRES, Alex Sander Xavier - Súmula Vinculante e Liberdades Fundamentais, p. 54

[375] BROWNSWORD, Roger - *What the World Needs Now: Techno-Regulation,Human Rights and Human Dignity*, p. 211.

[376] BROWNSWORD, Roger - *Op. cit.*, p. 211.

[377] *Id., ibdem*.

No mesmo sentido, também pode ser lembrada a posição de KOOPS, para quem a "aceitabilidade" consistiria em aspecto fundamental da qualquer forma de regulação. Conforme o jurista, a redução ou substituição do "dever" ("*ought*") ou "não dever" ("*ought not*") por "poder" ("*can*") ou "não poder" ("*cannot*") ameaçaria a flexibilidade e interpretação humana das normas, que seriam elementos essenciais da prática do Direito[378].

Em uma "comunidade moral", as pessoas não agiriam de acordo com os preceitos normativos por serem constrangidos a isso, mas sobretudo por subscreverem, aderirem racional e livremente a tais regras[379].

A liberdade seria, em conclusão, um componente indispensável para se atribuir valor moral à conduta humana.

Seja como for, do ponto de vista do aprendizado civilizatório ou do desenvolvimento da sensibilidade moral, a imposição coercitiva parece encerrar uma metodologia de baixa eficiência didática no longo prazo.

Ilustrativo, nesse contexto, é o exemplo fornecido por YEUNG quanto às formas de controle de acesso nas estações de metrô[380].

Em cidades como Paris, as barreiras impostas (do chão ao teto) obstariam totalmente a entrada de pessoas sem os respectivos bilhetes. Essa modalidade de controle findaria por enfraquecer o "auto-controle" ou "auto-contenção moral" ("*moral self-restraint*") e, por consequência, a "demoralização" ("*de-moralising effect*") do comportamento[381]. O resultado seria uma espécie de "flacidez" ou mesmo "atrofia" da "musculatura" moral. Ademais, os efeitos deletérios decorreriam da produção de uma certo embrutecimento moral ante a ausência de uma adesão livre e consciente à norma e por privar da ponderação racional acerca do sentido da diretriz normativa.

Em contraste, o efeito oposto poderia ser constatado nos lugares (v.g. Londres) nos quais referidos obstáculos são parciais ("*waist-high ticket barriers*"), não impedindo totalmente a infração da regra. Em referidas cidades, constatar-se-ia o desenvolvimento de um maior refinamento moral, diante do fato de os usuários do sistema respeitarem os comandos regulatórios por força de uma conduta espontânea[382].

---

[378] *Apud* LEENES, Ronald E - *Op. cit.*, p. 150.
[379] Nas palavras de BROWNSWORD: "*Ideal-typically, the fully moral action will involve an agent doing the right things (as a matter of act morality) for the right reasons (as a matter of agent morality)*"*Apud* LEENES, Ronald E - *Op. cit.*, p. 150.
[380] *Apud* LEENES, Ronald E - *Op. cit.*, p. 160.
[381] *Id., ibidem*.
[382] *Apud* LEENES, Ronald E - *Op. cit.*, p. 160.

Evidentemente, cumpriria ponderar que, a depender do bem jurídico protegido talvez se pudesse cogitar da razoabilidade da existência de instrumentos tecno-regulatórios com elementos de contenção absoluta do comportamento ilícito. O direito à vida vem a lume como um exemplo quase inconteste de um valor que poderia atrair uma proteção desse jaez[383].

3.1.5.4. Breve avaliação sobre a abordagem tecnoregulatoria

A modulação da conduta humana figura no epicentro dos principais desafios relativos às possibilidade e limites de atuação do Poder Legislativo e de outras instâncias regulatórias do Estado de Direito Democrático.

O advento de sofisticados artefatos tecnológicos que permitem, sobretudo com o suporte da inteligência artificial, a otimização da eficácia normativa dos regramentos estatais parece sinalizar para o atingimento de patamares civilizatórios nunca alcançados, ao inscrever em tais instrumentos mecanismos intrínsecos de regulação de conduta.

Contudo, como visto, ao lado das possibilidades de supervisão e imposição eficiente da conformação do comportamento humanos às regras objeto de regular processo legislativo, as novas ferramentas tecno-regulativas artificialmente inteligentes podem comprometer direitos e valores essenciais.

Impõe-se, nesse cenário, promover uma releitura do princípio da legalidade, com o objetivo de assegurar a preservação das "liberdades fundamentais" na dinâmica da regulação por meio da tecnologia.

A inserção de definições arquiteturais apriorísticas no processo de criação e desenvolvimento de sistemas tecno-normativos alinhados com os direitos humanos ("*human-rights-by-design*") afigura-se como uma das soluções para alcançar tal desiderato.

[383] Em um exercício de mera especulação teórica e um tanto pueril apenas para ilustrar o raciocínio, poder-se-ia imaginar o caso de uma trava eletrônica nas armas de fogo que impediria o disparo fora das hipóteses excepcionais autorizadas por lei (v.g. legítima defesa).

3.2. Cenário comparado

3.2.1. União Europeia[384]

Em abril de 2021, a Comissão Europeia propôs novas regras destinadas a garantir a confiabilidade dos sistemas de IA[385].

De acordo com a descrição formulada pela Comissão referida, "a conjunção do primeiro quadro jurídico em matéria de inteligência artificial e de um novo Plano Coordenado com os Estados-Membros garantirá a segurança e a defesa dos direitos fundamentais das pessoas e das empresas, reforçando simultaneamente o investimento, a inovação e a utilização da inteligência artificial, em toda a UE"[386].

Um dos motivos centrais que conduziram a essa iniciativa regulatória concerne à necessidade de se assegurar confiança na utilização de soluções tecnológicas artificialmente inteligentes, sobretudo em virtude dos riscos de malferimento de direitos e garantias fundamentais[387].

Do mesmo modo que se deu quanto às políticas normativas relativas à proteção de dados pessoais[388], a União Europeia pretende, assim, assumir algum protagonismo no processo de regulação da IA.

---

[384] Vale notar que grande parte das considerações lançadas no item "3.2.1. União Europeia" foi extraída do seguinte texto, de autoria própria: PAULA, Gáudio Ribeiro de - A Proposta Europeia para regulação da inteligência artificial.

[385] O anúncio foi feito em 21 de abril do ano em curso (2021). Fonte: https://ec.europa.eu/commission/presscorner/detail/pt/ip_21_1682. Acesso em 21/04/21.

[386] Id., ibid.

[387] MARGRETHE VESTAGER, Vice-presidente executiva da entidade "*Uma Europa Preparada para a Era Digital*", ressaltou, nesse sentido, que: "No domínio da inteligência artificial, a confiança é um imperativo, não um acessório. Com esta regulamentação histórica, a UE lidera o desenvolvimento de novas normas mundiais, para garantir uma inteligência artificial de confiança. Ao estabelecermos as normas, podemos abrir caminho à tecnologia ética em todo o mundo e garantir simultaneamente que a UE se mantenha competitiva. Preparadas para o futuro e favoráveis à inovação, as nossas regras serão aplicadas quando estritamente necessário: sempre que se trate da segurança e dos direitos fundamentais dos cidadãos da UE". Fonte: https://ec.europa.eu/commission/presscorner/detail/pt/ip_21_1682. Acesso em 21/04/21.

[388] Como se recorda, a "*General Data Protection Regulation*" (GDPR) – aprovada por meio do Regulamento 2016/679 – inspirou diversas iniciativas semelhantes, entre as quais a Lei Geral de Proteção de Dados Pessoais (LGPD) brasileira – Lei 13.709/18.

Não se pode deixar de destacar, contudo, que já existem diversas propostas de enfrentamento da questão, seja por meio de mecanismos convencionais de "*hard law*"[389], seja por meio de instrumentos de "*soft law*"[390].

Portanto, o projeto europeu não seria, propriamente, inédito, mas exerce um peso sistêmico significativo e pode induzir outros Estados, blocos político-econômicos e organizações internacionais a encamparem projetos semelhantes.

3.2.1.1. Breve escorço histórico

3.2.1.1.1. Recomendações do Parlamento Europeu à Comissão sobre disposições de Direito Civil sobre Robótica

Um dos primeiros esforços mais relevantes empreendidos no âmbito europeu foi o conjunto de Recomendações do Parlamento Europeu à Comissão sobre disposições de Direito Civil sobre Robótica, aprovadas em 16 de fevereiro de 2017.

Em seus *consideranda*, depreende-se um notável receio no concernente aos riscos da ausência de regulação quanto às várias manifestações da IA, que poderiam levar a uma nova revolução industrial[391].

---

[389] Nesse âmbito as iniciativas são escassas, consoante as informações contidas em relatório das Nações Unidas sobre a matéria. Organização das Nações Unidas (ONU) - *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, p. 16. De modo geral, além de diretrizes programáticas a conduzirem políticas públicas e definirem estratégias econômicas gerais, as tentativas regulatórias têm por objeto aspectos específicos relativos ao desenvolvimento ou implementação de sistemas de IA em determinados segmentos (v.g. armas letais e veículos autônomos). Nos Estados Unidos, dos 39 projetos legislativos sobre o assunto, apenas quatro foram convertidos em lei, conforme levantamento realizado pela Biblioteca do Congresso Norte-Americano, em 2019 – Fonte: Regulation of Artificial Intelligence: The Americas and the Caribbean, disponível em https://www.loc.gov/law/help/artificial-intelligence/americas.php [Acesso em 21/04/21]

[390] Podem ser citados, dentre outros, os já mencionados: "Carta sobre Ética Robótica" da Coreia do Sul, de 2007; "Iniciativa Global do Instituto de Engenharia Elétrica e Eletrônica (IEEE) sobre Ética de Sistemas Autonomos e Inteligentes" ("*IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*"); Declaração de Princípios de Asilomar; a Recomendação da OCDE sobre inteligência artificial; e declarações de princípios sobre IA do Google, Microsoft e IBM.

[391] Dentre os que sobressaem, podem ser reproduzidos os seguintes *consideranda* "A. Considerando que desde o Frankenstein de Mary Shelley ao mito clássico do Pigmaleão, passando pela história do Golem de Praga ao robô de Karel Čapek, que cunhou o termo, as pessoas têm fantasiado acerca da possibilidade de construir máquinas inteligentes, frequentemente androides com características humanas; B. Considerando que, agora que a humanidade se encontra no limiar de uma era em que robôs, «bots», androides e outras manifestações de inteligência artificial («IA») cada vez mais sofisticadas parecem estar preparadas para desencadear uma nova revolução industrial, que provavelmente não deixará nenhuma camada da sociedade intacta, é extremamente importante que a legislatura pondere as suas implicações e efeitos a nível jurídico e ético, sem colocar entraves à inovação; C. Considerando que é necessário criar uma definição geralmente aceite de robô e de IA que seja flexível e não crie obstáculos à inovação; [...]". PARLAMENTO EUROPEU - Recomendações do à Comissão sobre disposições de Direito Civil sobre Robótica.

O Parlamento Europeu sugeriu a introdução no mercado interno da União de um sistema abrangente de registro de robôs avançados, sempre quanto a certas categorias específicas de robôs. Além disso, instou a Comissão a estabelecer critérios para a classificação de robôs para fins de registro, assim como a estudar a possibilidade de criar uma Agência designada da UE para a robótica e a inteligência artificial para esse e outros fins[392].

Salientou, de outro lado, que o desenvolvimento das tecnologias da robótica deveria ser orientado para "complementar as capacidades humanas", e não para as substituir. Ressaltou, ainda, a essencialidade de se preservar o controle humano sobre as máquinas. Curiosamente, chamou a atenção para os riscos de virtual "ligação emocional entre seres humanos e robôs", sobretudo em relação a grupos vulneráveis (e.g. crianças, idosos e pessoas com deficiência)[393].

Em relação à abrangência regulatória, recomendou uma abordagem a nível da UE, de modo a obstar a "fragmentação do mercado interno", mas sublinhou, outrossim, a relevância do "princípio do reconhecimento mútuo na utilização transfronteiriça de robôs e de sistemas robóticos" [394].

Ao abordar os princípios éticos, observou as tensões a serem tomadas a sério quanto aos riscos concernentes a "segurança, saúde e proteção, liberdade, privacidade, integridade e dignidade, autodeterminação, não discriminação e proteção dos dados pessoais de seres humanos"[395].

Nesse mesmo ponto, divisou a necessidade de atualização e complementação do quadro jurídico então em vigor, a partir de um repositório de fundamentos éticos "orientador, claro, rigoroso e eficiente". Propõe, nessa esteira, a redação de um "código de conduta para os engenheiros de robótica" (o que também se aplicaria aos atores de IA) [396].

Realçou, também relativamente aos aspectos éticos, o princípio da transparência, de forma a garantir a explicitação inteligível dos fundamentos que sustentem "qualquer decisão tomada com recurso a inteligência artificial que possa ter um impacto substancial sobre a vida de uma ou mais pessoas". A esse propósito, idealizou a introdução de "*caixas negras*" (ou "caixas pretas") nos robôs avançados, para preservar um "*log*" (registro) intangível dos

[392] PARLAMENTO EUROPEU - Recomendações do à Comissão sobre disposições de Direito Civil sobre Robótica.
[393] PARLAMENTO EUROPEU – *Op. cit.*
[394] PARLAMENTO EUROPEU – *Op. cit.*
[395] PARLAMENTO EUROPEU – *Op. cit.*
[396] PARLAMENTO EUROPEU – *Op. cit.*

"dados relativos a todas as operações realizadas pela máquina, incluindo os passos da lógica que conduziu à formulação de eventuais decisões" [397].

Fez expressa alusão, também, aos princípios da "beneficência, não-maleficência, autonomia e justiça", assim como aos princípios e valores positivados no art.º 2.º do Tratado da União Europeia e na Carta dos Direitos Fundamentais, entre os quais indicou, expressamente, "a dignidade do ser humano, a igualdade, a justiça e a equidade, a não-discriminação, o consentimento esclarecido, o respeito da vida privada e familiar e a proteção de dados". Referiu, por fim, a "outros princípios e valores subjacentes do direito da União", tais como a "não estigmatização, a transparência, a autonomia, a responsabilidade individual e a responsabilidade social, e em códigos e práticas éticas existentes" [398].

Até aqui, a Comissão ainda não teria incorporado a integralidade desse amplo corpo de recomendações do Parlamento Europeu[399].

3.2.1.1.2. Comunicação da Comissão Europeia sobre IA para a Europa e Livro Branco sobre IA

Entretanto, em abril de 2018, a Comissão Europeia publicou relatório contendo sua estratégia na abordagem da IA[400]. Iniciou por destacar que União Europeia (UE) deve ter uma abordagem coordenada para aproveitar ao máximo as oportunidades oferecidas pela IA e enfrentar os novos desafios por ela trazidos[401].

Contrariamente às intenções manifestadas por diversos membros do Parlamento Europeu, a Comissão não havia proposto, contudo, qualquer medida regulatória da IA, naquela altura[402].

Ao invés disso, comprometera-se a desenvolver um conjunto de diretrizes que veio a ser publicado no final de 2018 sob o título "Plano de Ação Coordenada sobre IA" (*"Coordinated Action Plan on AI"*) que definiu os objetivos e projetos para uma estratégia europeia mais abrangente para enfrentamento dos desafios emergentes das novas tecnologias relativas a IA[403].

---

[397] PARLAMENTO EUROPEU – *Op. cit.*
[398] PARLAMENTO EUROPEU – *Op. cit.*
[399] Informação disponível em: https://europa.eu/rapid/press-release_IP-18-3362_pt.htm
[400] Comissão Europeia – *Artificial Intelligence for Europe*.
[401] Comissão Europeia – *Artificial Intelligence for Europe*, p. 3.
[402] MARCHANT, Gary – *Op. cit.*, p. 10.
[403] *Id. ibid.*

A Comissão ressaltou que, embora a auto-regulação possa fornecer um primeiro conjunto de parâmetros ("*benchmarks*") a partir dos quais sistemas emergentes podem ser comparados e avaliados, as autoridades públicas não podem olvidar a necessidade de formar estruturas regulatórias para alinhar o progresso da IA aos valores e direitos fundamentais[404].

Nesse cenário, afirmou que acompanharia os desenvolvimentos do fenômeno e, se necessário, revisaria os quadros jurídicos existentes para melhor adaptá-los a desafios específicos, em particular para garantir o respeito dos valores básicos e dos direitos abraçados pela União Europeia[405].

É aqui que a abordagem sustentável da UE em relação às tecnologias criaria uma vantagem competitiva, ao firmar os valores consensuais de seus membros como centro de gravidade axiológico em suas intervenções para direcionar as mudanças tecnológicas[406].

De acordo com a Comissão, a UE deveria, por conseguinte, garantir que a IA fosse desenvolvida e aplicada num quadro jurídico-normativo adequado, de forma a promover a inovação, de um lado, e, de outro, o respeito aos valores e direitos fundamentais, assim como aos princípios éticos[407].

Destacava-se o recurso a "*regulatory sandboxes*", caixas de areia regulatórias (em uma tradução literal), campos de teste para avaliar novos modelos regulatórios em modelos de negócio ainda pendentes de regulação. A Comissão propusera sua utilização nos entroncamentos de inovação digital ("*Digital Innovation Hubs*"), em áreas diversas (v.g. transporte, agricultura, saúde, etc.) nas quais venham a surgir tecnologias disruptivas[408].

Após a publicação, em 2018, da Estratégia Europeia para a inteligência artificial, o Grupo de Peritos de Alto Nível em Inteligência Artificial (GPAN) elaborou orientações para uma inteligência artificial confiável, em 2019, assim como uma lista de avaliação, em 2020[409].

Em 2020, publicou-se o "Livro Branco da Comissão sobre a inteligência artificial". O documento definiu visão europeia sobre o tema: "um ecossistema de excelência e confiança", que pavimentou o caminho para a proposta agora apresentada[410].

---

[404] Comissão Europeia – *Op. cit.*, p. 15.
[405] MARCHANT, Gary – *Op. cit.*, p. 10.
[406] Comissão Europeia – *Op. cit.*, p. 2.
[407] *Id., ibid.*
[408] Comissão Europeia – *Op. cit.*, p. 9.
[409] Fonte: https://ec.europa.eu/commission/presscorner/detail/pt/ip_21_1682. Acesso em 21/04/21.
[410] "A consulta pública a respeito do Livro Branco sobre a inteligência artificial granjeou ampla participação mundial. O «Relatório sobre as implicações em matéria de segurança e de responsabilidade decorrentes da inteligência artificial, da Internet das coisas e da robótica», que acompanha o Livro Branco, concluiu que a atual legislação sobre segurança dos produtos contém uma série de lacunas, a abordar, nomeadamente, na Diretiva Máquinas". Fonte: https://ec.europa.eu/commission/presscorner/detail/pt/ip_21_1682. Acesso em 21/04/21.

3.2.1.1.3. Carta ética europeia sobre o uso da inteligência artificial em sistemas judiciais e seu ambiente

Dentro do contexto europeu, pode-se mencionar, ainda e em paralelo, por sua relevância jurídica, a Carta ética europeia sobre o uso da inteligência artificial em sistemas judiciais e seu ambiente, elaborada pela Comissão Europeia para Eficiência da Justiça[411].

Adotada na sua 31ª sessão plenária em Estrasburgo, nos dias 3 e 4 de dezembro de 2018, o documento enuncia 5 princípios: i) respeito aos direitos fundamentais ("*Principle of respect for fundamental rights*") – para garantir que o design e a implementação das ferramentas de IA sejam compatíveis com os direitos; ii) não discriminação ("*Principle of non-discrimination*") – para prevenir o surgimento ou intensificação de qualquer forma de discriminação relativamente a indivíduos ou coletividades; iii) qualidade e segurança – ("*Principle of quality and security*") – no processamento de decisões judiciais, deve-se utilizar fontes seguras de dados em um ambiente tecnológico confiável; iv) transparência, imparcialidade e justiça – ("*Principle of transparency, impartiality and fairness*") – de modo a tornar os métodos de processamento de dados acessíveis, inteligíveis e passíveis de submissão a auditorias externas; v) controle do usuário ("*Principle under user control*") – garantir que os usuários mantenham o controle sobre suas escolhas[412].

Muito embora a Carta dirija-se mais específica e explicitamente aos órgãos jurisdicionais, vários de seus princípios (a maior parte, em verdade) revelam-se aplicáveis em outros âmbitos, coincidindo com alguns dos contidos na proposta legislativa em exame.

3.2.1.2. Proposta de regulação europeia

3.2.1.2.1. Objetivos

Ante a necessidade de manter a coerência intranormativa e o alinhamento com as políticas já adotadas no âmbito europeu, a Comissão enunciou os seguintes objetivos da proposta regulatória: a) garantir que os sistemas de IA sejam seguros e respeitem aos direitos

[411] "*European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*", no original em inglês. Sua íntegra encontra-se disponível em https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment [Consult. 3 abr. 2019].
[412] CEPEJ - *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*.

fundamentais e valores albergados pela UE; b) preservar a segurança jurídica para facilitar o investimento e a inovação em IA; c) melhorar a governança e assegurar a eficácia dos direitos fundamentais, assim como o controle dos requisitos de segurança aplicáveis aos sistemas de IA; d) facilitar o desenvolvimento de um ambiente econômico único (no contexto do mercado comum) para aplicações de IA compatíveis com as regras jurídicas, seguras e confiáveis[413].

A tentativa de regulação segue uma abordagem baseada na hierarquização do risco criado pelos sistemas de IA, de modo a considerar três níveis: i) inaceitável – quanto ao qual seriam inteiramente vedadas as práticas; ii) alto – em que haveria severas restrições ao desenvolvimento, implementação e uso; e iii) baixo ou mínimo – no qual haveria tolerância quase plena, sem justificar qualquer intervenção restritiva particular[414].

3.2.1.2.2. Estrutura

O projeto normativo encontra-se decomposto em doze títulos: i) disposições gerais; ii) práticas de IA vedadas; iii) sistemas de IA de alto risco; iv) obrigações de transparência quanto a determinados sistemas de IA; v) medidas de estímulo à inovação; vi) governança; vii) banco de dados europeu de sistemas de IA de alto risco; viii) monitoramento "pós-comercialização", compartilhamento de informações, e "vigilância" do mercado; ix) códigos de conduta; x) confidencialidade e sanções; xi) delegações de poderes e procedimento do comitê; e xii) disposições finais[415].

O espectro regulatório, como se vê, é bastante abrangente e audacioso. Resulta de estudo multidisciplinar consistente e aprofundado de diversos aspectos concernentes aos impactos jurídicos da criação e emprego de sistemas de IA. Foi fruto de inúmeras contribuições de diversos "*stakeholders*"[416].

Passa-se, em seguida, a uma rápida apresentação dos pontos que mais sobressaem, em uma primeira análise.

---

[413] Tradução livre do original em inglês. O texto completo pode ser encontrado em: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence [Acesso em 21/04/21]

[414] É o caso da maior parte das aplicações de IA disponíveis, atualmente (v.g. jogos, filtros de spam, digitação preditiva e reconhecimento de voz).

[415] O texto completo pode ser encontrado em: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence [Acesso em 21/04/21]

[416] Foram recebidas 1.215 sugestões, advindas de empresas, associações, entidades da sociedade civil, instituições acadêmicas, autoridades públicas e indivíduos. Fonte: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence [Acesso em 21/04/21]

3.2.1.2.3. Disposições gerais – escopo e definições

O primeiro título define a abrangência normativa das regras propostas, além de delinear os principais conceitos.

No tocante ao escopo "objetivo", vale ressaltar que a regulação pretendida abarcaria disponibilização e utilização no mercado europeu (UE) de aplicações de IA, independentemente de onde os seus desenvolvedores ou "provedores" advenham (art. 2º, 1, 'a'). Estariam excetuados os sistemas desenvolvidos para fins militares, os quais não seriam abrangidos pelas regras "*in fieri*" (art. 2º, 3).

De outro lado, sob a perspectiva "subjetiva", as novas disposições alcançariam todos os usuários localizados em território dos países que integram a UE (art. 2º, 1, 'b'). Não seriam alcançadas, contudo, as autoridades públicas de Estados estrangeiros e organizações internacionais, em dadas condições (art. 2º, 4).

Entre as 44 definições apresentadas, por evidenciarem alguns dos pontos sensíveis enfrentados, merecem destaque as seguintes: i) sistema de inteligência artificial – programa desenvolvido por meio de quaisquer das técnicas ou abordagens listadas no Anexo I (v.g. "*machine learning*"), para um determinado conjunto de objetivos definidos pelo homem, e que geram resultados como informações, previsões, recomendações ou decisões que influenciam os ambientes com os quais interagem; ii) provedor – pessoa física ou jurídica, autoridade pública, agência ou outro órgão que desenvolve ou possui sistema de IA com vista a colocá-lo no mercado ou em serviço em seu próprio nome ou marca registrada, de forma onerosa ou gratuita; iii) usuário – pessoa física ou jurídica, autoridade pública, agência ou outro órgão que utilize sistema de IA sob sua autoridade, exceto quando usado em atividade não profissional; iv) uso indevido razoavelmente previsível ("*reasonably foreseeable misuse*") – uso em desacordo com a finalidade pretendida, mas que pode resultar de comportamento humano previsível ou interação com outros sistemas; v) sistema de reconhecimento de emoção ("*emotion recognition system*") – sistema criado com a finalidade de identificar ou inferir emoções ou intenções de pessoas físicas com base em seus dados biométricos coletados; vi) sistema de categorização biométrica ("*biometric categorisation system*") – sistema desenvolvido com o propósito de associar pessoas físicas a categorias específicas, como sexo, idade, cor do cabelo, cor dos olhos, tatuagens, origem étnica, orientação sexual ou política, a partir da coleta de dados biométricos; e vii) incidente sério ('serious incident') – que, direta ou indiretamente, possa conduzir à morte de uma pessoa ou causar sérios danos à sua integridade física, à propriedade ou ao meio ambiente, assim como

levar a uma grave e irreversível perturbação na gestão ou operação de infraestrutura crítica (art. 3º)**.**

3.2.1.2.4. Práticas de IA vedadas

O avanço legislativo mais relevante e em linha com várias ponderações de estudiosos e entidades da sociedade civil talvez seja a vedação de diversas práticas relativas à inteligência artificial, em virtude dos riscos inaceitáveis de vulneração de direitos fundamentais.

Estariam incluídos no rol de proibições aquelas práticas destinadas a induzir ou modificar o comportamento de uma pessoa, de modo a poder causar algum dano físico ou psicológico, tais como: a) o emprego de técnicas subliminares, não detectadas pela consciência de alguém; e b) a exploração de vulnerabilidades de um grupo específico de pessoas, resultantes da idade, deficiência física ou mental (art. 5º, 1, 'a' e 'b').

Seria vedada, de outro lado, a avaliação ou classificação da confiabilidade de pessoas físicas, a partir de seu comportamento social ou características pessoais ou de personalidade conhecidas ou previstas, com a "pontuação social" de que decorra tratamento desfavorável de certas pessoas ou grupos: i) em contextos sociais que não estão relacionados com os contextos em quais os dados originalmente gerados ou coletados; ou ii) que seja injustificado ou desproporcional ao comportamento identificado ou sua gravidade (art. 5º, 1, 'c').

Trata-se de uma espécie de repúdio ao modelo adotado no conhecido e já referido "Sistema de Crédito Social" (SCS) chinês, o qual, recorde-se, consiste em projeto de engenharia social proposto para modelar o comportamento individual, por meio de uma sistemática de pontuações destinada aos seus 1,4 bilhão de cidadãos, premiando os merecedores de confiança punindo os desobedientes[417]. As bonificações levariam em conta atitudes socialmente louváveis (v.g. trabalho voluntário e doação de sangue), hábitos de consumo, conduta no trânsito, respeito às diretrizes do Partido Comunista e a ordens judiciais, entre outros critérios[418]. Com lastro em tais elementos e conforme relatório da Comissão Nacional de Desenvolvimento e Reforma do país, estima-se que cerca de 7 milhões de chineses teriam sido impedidas de embarcar em vôos comerciais por serem reputados

[417] XU, V. X., & XIAO, B. - *China's social credit system seeks to assign citizens scores, engineer social behaviour*.
[418] XU, V. X., & XIAO, B. - *Op. Cit.*

"indignos de confiança", assim como outros 3 milhões teriam deixado de viajar em trens de alta velocidade pelo mesmo motivo[419].

Ademais, de acordo com a normativa em vias de implementação, também estaria banido o uso de sistemas de identificação biométrica remota em tempo real e em lugares públicos.

Haveria, entretanto, duas ressalvas a autorizarem o seu emprego: i) a busca de vítimas potenciais de crime, assim como de crianças desaparecidas; ii) a prevenção de ameaça concreta e iminente à vida ou integridade física, o que incluiria ataques terroristas; iii) a detecção, localização, identificação de autor ou suspeito de determinadas infrações penais (art. 5º, 1, 'd').

Os critérios para avaliar a licitude quanto ao uso de tais sistemas seriam: a) a natureza da situação, além da seriedade, probabilidade e gravidade do dano resultante da não utilização da identificação biométrica; b) os impactos nos direitos e liberdades das pessoas envolvidas (art. 5º, 2).

3.2.1.2.5. Sistemas de IA de alto risco

De acordo com a relação contida nos Anexos II e III referidos na proposta normativa (art. 6º), são considerados de elevado risco os sistemas utilizados em, dentre outros: i) infraestruturas críticas que possam comprometer a vida ou integridade física (v.g. transportes); ii) educação ou formação profissional que possam restringir o acesso à educação e a evolução profissional de alguém (v.g. classificação de exames); iii) componentes de segurança de produtos (v.g. cirurgia assistida por robôs); iv) emprego, gestão de trabalhadores e acesso ao trabalho por conta própria (v.g. análise de currículo em processos seletivos); v) serviços públicos e privados essenciais (v.g. pontuação de crédito para obtenção de empréstimos); vi) "aplicação coerciva da lei" que possa interferir com os direitos fundamentais das pessoas (v.g. "avaliação da fiabilidade de provas"); vii) gestão da migração e do controle de fronteiras (v.g. verificação da autenticidade de documentos de viagem); e viii) administração da justiça e processos democráticos (v.g. "aplicação da lei a um conjunto específico de fatos")[420].

Diversamente dos sistemas quanto aos quais se identifica situação de risco inaceitável, aqueles reputados de alto risco não são vedados, mas sujeitos a severas restrições.

[419] XU, V. X., & XIAO, B. - *Op. Cit.*
[420] Fonte: https://ec.europa.eu/commission/presscorner/detail/pt/ip_21_1682 [Acesso em 21/04/21]

É o caso da necessidade de implementação de sistema de gestão de risco, mediante processo iterativo contínuo executado ao longo de todo o ciclo de vida do sistema (art. 9º).

Também seria obrigatória a manutenção de programa de governança de dados, mediante o uso de técnicas envolvendo o treinamento de modelos com dados, submetidos a validação e teste (art. 10).

Deveria ser produzida, ainda, exaustiva documentação técnica, antes de o sistema entrar em operação, a ser mantida atualizada (art. 11), assim como a rigorosa manutenção de registros, por meio de recursos que permitam a gravação automática de eventos ("logs"), de forma a permitir a rastreabilidade de eventuais erros (art. 12).

Os sistemas de alto risco deveriam, outrossim, ser projetados de forma a garantir que sua operação é suficientemente transparente para permitir que os usuários interpretem os dados gerados. Deveriam ser acompanhados por instruções de uso concisas, completas, acessíveis e compreensíveis para os usuários (art. 13).

A supervisão humana deveria ser assegurada, com ferramentas de "interface homem-máquina" apropriadas, de modo a prevenir ou minimizar os riscos para a saúde, segurança ou direitos fundamentais (art. 14).

Por fim, exigir-se-iam de tais sistemas com risco acentuado os atributos de precisão, robustez e segurança em todo seu ciclo de vida, assegurando a "resiliência" a erros, falhas ou inconsistências que podem ocorrer dentro dos sistemas ou ambientes em que operam, em particular devido à interação com pessoas físicas ou outros sistemas (art. 15).

3.2.1.2.6. Transparência

Nas situações de risco limitado, as principais exigências impostas referem-se à compulsoriedade da transferência quanto a diversos dados e informações[421].

Os sistemas destinados a interagir com pessoas devem ser projetados de tal forma que fique clara a natureza artificial do "interlocutor" ou da interface, a menos que isso seja óbvio, ante as circunstâncias ou contexto de uso (art. 52, 1)[422]. A diretriz não se aplica, contudo, a

[421] Fonte: https://ec.europa.eu/commission/presscorner/detail/pt/ip_21_1682 [Acesso em 21/04/21]

[422] Com o exponencial progresso nos sistemas de interação dotados de IA, torna-se cada vez mais árduo distinguir a condição humana do interlocutor. Estariam em vias de superar o conhecido teste de Turing. A esse respeito, um exemplo ilustrativo é o "Google Duplex", que permitiria o agendamento de reservas em restaurantes e outros estabelecimentos comerciais por meio da aplicação contida no smartphone. A esse propósito, também se instalou nos EUA debate em torno da necessidade de haver a identificação do agente como artificialmente inteligente. Vide matérias disponíveis em: https://towardsdatascience.com/did-google-duplex-beat-the-turing-test-yes-and-no-a2b87d1c9f58 e https://www.mercurynews.com/2018/05/18/googles-human-like-speaking-bot-may-run-into-legal-issues-say-experts/ [Acesso em 21/04/21]

sistemas autorizados por lei a detectar, prevenir, investigar e processar infrações criminais (art. 52, 1).

Na mesma esteira, as pessoas expostas a sistemas de reconhecimento de emoção ou de categorização biométrica devem ser informadas acerca de seu funcionamento (art. 52, 2). Ressalva-se, igualmente, o uso permitido por lei para detectar, prevenir e investigar infrações criminais (art. 52, 2).

Os usuários de sistemas que permitam a manipulação de imagem, áudio ou vídeo, de modo a tornar tais arquivos indistinguíveis dos autênticos ou originais ("*deep fake*"), devem indicar a alteração artificial ao propagarem o conteúdo (art. 52, 3). No entanto, não se aplica quando sua utilização é "autorizada por lei a detectar, prevenir, investigar e julgar infrações penais ou for necessário para o exercício do direito à liberdade de expressão e direito à liberdade das artes e ciências garantidas na Carta dos Direitos Fundamentais da UE, e sujeitas a salvaguardas adequadas dos direitos e liberdades de terceiros" (art. 52, 3).

3.2.1.3. Breve avaliação inicial da abordagem europeia

O Parlamento Europeu e os Estados-Membros ainda devem avaliar as propostas da Comissão relativas a essa "abordagem europeia à inteligência artificial e às máquinas". Se vierem a ser acolhidas, como se sabe, os regulamentos serão diretamente aplicáveis em toda a UE[423].

Entre os fins almejados, encontra-se a tentativa de a Europa alcançar a posição de liderança como "*global rulemaker*"[424] quanto ao tema, a partir de uma perspectiva "centrada no ser humano, sustentável, segura, inclusiva e fiável"[425].

Cuida-se de uma prioridade na agenda conduzida pelo braço executivo da UE, de acordo com as posições externadas por URSULA VON DER LEYEN, atual presidente da Comissão Europeia[426].

A esperada celeridade na implementação das medidas deverá obrigar os grandes "*players*" do Vale do Silício e de outros "*clusters*" tecnológicos a ajustarem algumas de suas práticas no curto prazo, tal como se deu no caso da proteção de dados (em virtude da entrada

[423] Fonte: https://ec.europa.eu/commission/presscorner/detail/pt/ip_21_1682 [Acesso em 21/04/21]
[424] Fonte: https://www.politico.eu/article/europe-throws-down-gauntlet-on-ai-with-new-rulebook/ [Acesso em 21/04/21]
[425] Fonte: https://ec.europa.eu/commission/presscorner/detail/pt/ip_21_1682 [Acesso em 21/04/21]
[426] Fonte: https://ec.europa.eu/commission/presscorner/detail/en/speech_20_294 [Acesso em 21/04/21]

em vigor da GPDR). Isso também pode representar um desafio à nova administração federal norte-americana[427].

A agilidade na construção do marco regulatório europeu atraiu algumas críticas quanto ao fato de ter sido algo benevolente com as companhias de tecnologia, além de ter franqueado certa abertura para o uso da IA para vigilância estatal[428].

Seja como for, são inegáveis os méritos da iniciativa da Comissão Europeia, ao endereçar uma das questões mais candentes do nosso tempo.

Nesse contexto, a formulação de modelos de regulação dos sistemas de inteligência artificial representa uma preocupação central em um cenário de profunda insegurança jurídica resultante da célere evolução experimentada sobretudo no domínio das novas tecnologias, com potenciais riscos importantes de vulneração a direitos humanos.

De outra parte, os novos recursos tecnológicos estão a amplificar o conhecido e antes mencionado problema do ritmo ("*pacing*"), relativo à dificuldade de a legislação acompanhar a rápida evolução dos fatos sociais[429]. O já evidente descompasso entre legislação e problemas jurídicos emergentes das novas tecnologias alcança no caso da IA um patamar inédito, ante o possível surgimento célere de realidades até aqui apenas imaginadas ou mesmo inconcebíveis[430].

Por conseguinte, ao descortinar algumas importantes possibilidades regulatórias, com consistência, abrangência e profundida, a Europa pode representar, de fato, um farol a

---

[427] A esse propósito, entre as primeiras avaliações a respeito da proposta regulatória europeia, eis o que MARK SCOTT pondera: "*While Washington has sought closer ties with Europe to counter China's growing tech ambitions, so far the U.S. hasn't followed the EU's lead on AI or on privacy. The new rules — which will now snake their way through Europe's legislative process — may widen the regulatory gulf between the two sides, even as Brussels pushes for closer coordination on its own tech priorities via a proposed Trade and Technology Council. At the same time, the proposed rules set the EU apart from China on tech. The fact that the rules have singled out social credit scoring — a tool used mainly in China — is a signal that Brussels wants to avoid uses of AI for authoritarian surveillance. 'It sends a clear message to China that the social credit system is incompatible with liberal democracies,' said Maroussia Lévesque, a researcher at the Berkman Klein Center at Harvard University. 'There is no room for mass surveillance in our society," said Commission Executive Vice President Margrethe Vestager'"*. Fonte: https://www.politico.eu/article/europe-throws-down-gauntlet-on-ai-with-new-rulebook/ [Acesso em 21/04/21]

[428] É o que ressalta MARK SCOTT: "*But that speed may come at the price of greater opposition to the fine print from civil society actors and EU lawmakers who must now parse the European Commission's proposal. Already, campaigners are voicing disappointment with a final Commission draft many of them say is too friendly to industry, and gives governments too wide a berth to use AI for surveillance"*. Fonte: https://www.politico.eu/article/europe-throws-down-gauntlet-on-ai-with-new-rulebook/ [Acesso em 21/04/21]

[429] HAGEMANN, Ryan & SKEES, Jennifer Huddleston & THIERER, Adam – *Soft Law For Hard Problems: The Governance Of Emerging Technologies In An Uncertain Future Intelligence*, p. 37.

[430] Veja-se, exemplificativamente, o já referido caso da imputação de responsabilidade ou mesmo personalidade jurídica a agentes artificialmente inteligentes, que vem sendo objeto de intenso debate doutrinário. A esse respeito: SOLUM, Lawrence B. - *Legal Personhood for Artificial Intelligences*; ANDRADE, Francisco & NOVAIS, Paulo & NEVES, José - *Issues on Intelligent Electronic Agents and Legal Relations; e* WETTIG, Steffen and ZEHENDNER, Eberhard – *A legal analysis of human and electronic agents*.

iluminar os caminhos a serem seguidos por outros órgãos legiferantes para compatibilizar inovação tecnológica e respeito aos direitos fundamentais.

Há se de registrar que, em 28 de setembro de 2022, foi apresentada proposta de diretiva 2022/0303 da Comissão Europeia a respeito da adaptação de regras de responsabilidade civil extracontratual para IA ("*AI Liability Directive*")[431].

3.2.2. China

A China tem disputado a primazia no desenvolvimento de tecnologias relacionadas à IA, em suas múltiplas frentes, v.g. acadêmicas, comerciais e mesmo regulatórias.

Em março de 2022, foi aprovada norma relevante dirigidas às empresas que exploram sistemas de recomendação "*online*" baseados em algoritmos[432].

De acordo com suas disposições, os provedores de serviços de recomendação algorítmica deveriam manter "orientações de valor", além de "disseminar vigorosamente energia positiva" e na "direção do bem"[433].

Por outro lado, deveriam se abster de promover atividades que prejudiquem a "segurança nacional" ou o "interesse social público", assim como que perturbem a "ordem econômica" ou a "ordem social" [434].

---

[431] Fonte: https://eur-lex.europa.eu/procedure/EN/2022_303 [Acesso em 29/09/22].

[432] Segue tradução para o inglês do texto original de seu art. 1º: "*Article 1: In order to standardize Internet information service algorithmic recommendation activities, carry forward the Socialist core value view, safeguard national security and the social and public interest, protect the lawful rights and interests of citizens, legal persons, and other organizations, and stimulate the healthy development of Internet information services; and on the basis of the 'Cybersecurity Law of the People's Republic of China', the 'Data Security Law of the People's Republic of China', 'the Personal Information Protection Law of the People's Republic of China', the 'Internet 'Information Service Management Rules," and other such laws and administrative regulations; these Provisions are formulated*". Fonte: https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022/ [Acesso em 10/09/22].

[433] Em seu art. 6º se lê: "*Article 6: Algorithmic recommendation service providers shall uphold mainstream value orientations, optimize algorithmic recommendation service mechanisms, vigorously disseminate positive energy, and advance the use of algorithms upwards and in the direction of good.* […]". Fonte: https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022/ [Acesso em 10/09/22].

[434] Ainda no art. 6º está previsto que: "*Article 6:* […] *Algorithmic recommendation service providers may not use algorithmic recommendation services to engage in activities harming national security and the social public interest, upsetting the economic order and social order, infringing the lawful rights and interests of other persons, and other such acts prohibited by laws and administrative regulations. They may not use algorithmic recommendation services to disseminate information prohibited by laws and administrative regulations, and shall take measures to prevent and curb the dissemination of harmful information*". Fonte: https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022/ [Acesso em 10/09/22].

O regulamento chinês obriga, ainda, as empresas de notificarem os usuários nas situações em que o algoritmo estiver desempenhando um papel na determinação de quais informações exibir para eles e dê aos usuários a opção de optar por não serem direcionados[435].

Tais prestadores de serviços de recomendação algorítmica devem, ademais, "proteger os direitos e interesses legítimos dos trabalhadores", concernentes, exemplificativamente, a remuneração do trabalho, descanso e férias, entre outros[436].

Se tais companhias venderem produtos ou prestarem serviços a consumidores, devem, outrossim, observar os "direitos comerciais justos dos consumidores", abstendo-se de conceder "tratamento diferenciado desarrazoado" nas condições comerciais, tais como preços de negociação, etc., e outras atividades ilícitas, com base nas tendências dos consumidores, hábitos comerciais e outras características semelhantes[437].

3.2.3. Estados Unidos

Nos Estados Unidos, reporta-se uma dupla fragmentação: i) a resultante das competências legiferantes autônomas atribuídas às unidades federativas; e ii) a concernente ao próprio objeto das normas existentes, que versam sobre diferentes aspectos particulares da IA[438].

---

[435] É o que estabelece o art.º 20º: "*Article 20: Where algorithmic recommendation service providers provide work dispatch services to workers, they shall protect workers' lawful rights and interests such as obtaining labor remuneration, rest and vacation, etc., and establish and perfect algorithms related to platform sign-on and allocation, remuneration composition and payment, work time, rewards, etc*". Fonte: https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022/ [Acesso em 10/09/22].

[436] De acordo com o art.º 21º: "*Article 21: Where algorithmic recommendation service providers sell products or provide services to consumers, they shall protect consumers' fair trading rights, they may not use algorithms to commit acts of extending unreasonably differentiated treatment in trading conditions such as trading prices, etc., and other such unlawful activities, on the basis of consumers' tendencies, trading habits and other such characteristics*". Fonte: https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022/ [Acesso em 10/09/22].

[437] Eis o que estabelecem os arts.º 16º e 17º: "*Article 16: Algorithmic recommendation service providers shall notify users in a clear manner about the situation of the algorithmic recommendation services they provide, and publicize the basic principles, purposes and motives, main operational mechanisms, etc., of the algorithmic recommendation services in a suitable manner. Article 17: Algorithmic recommendation service providers shall provide users with a choice to not target their individual characteristics, or provide users with a convenient option to switch off algorithmic recommendation services. Where users choose to switch off algorithmic recommendation services, the algorithmic recommendation service provider shall immediately cease providing related services. Algorithmic recommendation service providers shall provide users with functions to choose or delete user tags used for algorithmic recommendation services aimed at their personal characteristics. Where algorithmic recommendation service providers use algorithms in a manner creating a major influence on users' rights and interests, they shall give an explanation and bear related liability according to the law*". Fonte: https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022/ [Acesso em 10/09/22].

[438] Fonte: https://www.natlawreview.com/article/ai-regulation-where-do-china-eu-and-us-stand-today [Acesso em 19/09/22].

Diversas normas em vigor se concentram na criação de comissões para determinar como as agências estaduais podem utilizar a tecnologia de IA e estudar os possíveis impactos da IA na força de trabalho e nos consumidores[439]. Entre as exceções, pode ser mencionada a lei aprovada pelo Conselho de Nova Iorque que regula o uso de ferramentas na contração de trabalhadores, prevendo coimas de $500 a $1,500 por eventuais violações[440].

Entretanto, algumas iniciativas legislativas estaduais ainda pendentes de aprovação avançam na direção de regular a responsabilidade e a transparência dos sistemas de IA[441].

Em âmbito nacional, o Congresso dos EUA promulgou a Lei da Iniciativa Nacional de IA em janeiro de 2021, de modo a instituir "uma estrutura abrangente para fortalecer e coordenar atividades de pesquisa, desenvolvimento, demonstração e educação de IA em todos os departamentos e agências dos EUA" [442].

Seu objetivo consistiria em contribuir para o impulsionamento de avanços no desenvolvimento de sistemas de IA confiáveis, em todos os setores, de modo a beneficiar a população norte-americana[443].

O governo federal estadunidense se compromete, com tal medida, a fornecer os recursos necessários para facilitar o crescimento do capital humano para promover o avanço das tecnologias baseadas em IA, por meio da mobilização de instituições acadêmicas, empresas, e entidades governamentais[444].

O aludido diploma legal instituiu comissões e "forças-tarefa" com o objetivo de implementar uma estratégia nacional de IA, com o apoio de órgãos estatais, tais como a Comissão Federal de Comércio ("*Federal Trade Commission – FTC*") e os Departamentos de Defesa, Agricultura, Educação, Saúde e Serviços Humanos[445].

Ainda pendente de aprovação em nível nacional, pode ser referida a "Lei de Responsabilidade Algorítmica", que foi apresentada no Congresso norte-americano em fevereiro de 2022.

---

[439] *Id., ibid*.
[440] Fonte: https://www.shrm.org/hr-today/news/all-things-work/pages/regulations-ahead-on-artificial-intelligence.aspx [Acesso em 19/09/22].
[441] *Id., ibid.*
[442] Fonte: https://www.congress.gov/bill/116th-congress/house-bill/6216 [Acesso em 30/09/22].
[443] No original: "*In order to help drive forward advances in trustworthy artificial intelligence across all sectors and to the benefit of all Americans, the Federal Government must provide sufficient resources and use its convening power to facilitate the growth of artificial intelligence human capital, research, and innovation capacity in academia and other nonprofit research organizations, companies of all sizes and across all sectors, and within the Federal Government*". Fonte: https://www.congress.gov/bill/116th-congress/house-bill/6216 [Acesso em 30/09/22].
[444] Fonte: https://www.congress.gov/bill/116th-congress/house-bill/6216 [Acesso em 30/09/22].
[445] Fonte: https://www.natlawreview.com/article/ai-regulation-where-do-china-eu-and-us-stand-today [Acesso em 19/09/22].

A proposta nortearia o esforço regulamentador da Comissão Federal de Comércio destinado a impor obrigações a entes privados para que realizem avaliações de impacto nos processos automatizados de tomada de decisão, executados por agentes artificialmente inteligentes[446].

A aludida comissão (*FTC*) tem sido bastante ativa quanto ao tema. Em abril de 2021 emitiu um memorando acerca de práticas discriminatórias resultantes do emprego de soluções de IA[447]. Já em junho de 2022, a agência indicou que enviará um aviso prévio de regulamentação preliminar para "garantir que a tomada de decisão algorítmica não resulte em discriminação prejudicial", mediante consulta pública para colher contribuições dos interessados e especialistas a respeito do tema[448].

A *FTC* encaminhou, ainda, relatório ao Congresso dos EUA para indicar possibilidades de empregado da IA como instrumento para combater ilícitos cometidos pela Internet, como fraudes online, "*deep fake*", ou mesma a comercialização de drogas proibidas. No "*paper*" ressalta, contudo, a suscetibilidade da tecnologia de produzir resultados imprecisos, tendenciosos ou mesmo discriminatórios[449].

3.2.4. Portugal

Em relação ao Direito português, vale referir a Carta Portuguesa de Direitos Humanos na Era Digital, aprovada pela Lei n.º 27 de 17 de maio de 2021[450]. Há quem uma certa incongruência terminológica no título do documento, diante da circunstância de que os direitos humanos seriam universais, razão pela qual "o Estado português deveria ter promovido uma conferência internacional e não um procedimento legislativo"[451].

---

[446] Eis o objetivo do projeto tal como consta da sua literalidade: "*To direct the Federal Trade Commission to require impact assessments of automated decision systems and augmented critical decision processes, and for other purposes*". Fonte: https://www.congress.gov/bill/117th-congress/house-bill/6580/text [Acesso em 30/09/22].

[447] Fonte: https://www.natlawreview.com/article/ai-regulation-where-do-china-eu-and-us-stand-today [Acesso em 19/09/22].

[448] *Id., ibid.*

[449] *Id., ibid.*

[450] Fonte: https://dre.pt/dre/legislacao-consolidada/lei/2021-164870244 [Consult. 15 dez. 2022]

[451] ALEXANDRINO, José Melo - Dez breves apontamentos sobre a Carta Portuguesa de Direitos Humanos na Era Digital, p. 1. O autor promove críticas diversas à Carta, entre as quais: i) a imposição de limitações à liberdade de expressão não lastreadas na Constituição; ii) a redundância quanto à referência a direitos fundamentais já positivados no plano doméstico e internacional; e iii) as colisões com o Direito da União Europeia e com a jurisprudência do Tribunal de Justiça da União Europeia quanto à proteção de dados.

Seja como for, o instrumento versa sobre os direitos em ambiente digital, de forma geral, tais como liberdade de expressão e criação, proteção contra a desinformação, reunião, associação, privacidade, neutralidade da rede, identidade, entre outros.

No tocante especificamente à IA, reserva um artigo próprio (art.º 9º) para tratar do tema[452].

Em primeiro lugar, remete-se à centralidade dos direitos fundamentais, para assegurar um "justo equilíbrio" entre princípios relevantes como o da explicabilidade, segurança, transparência e responsabilidade. Menciona-se, de modo expresso, a tutela antidiscriminatória.

De outro lado, são impostas exigências para a utilização da IA como suporte a decisões "com impacto significativo na esfera dos destinatários". Nesses casos seria necessário observar os deveres de: i) comunicação aos interessados; ii) oferecimento de elementos que permitissem a auditagem do sistema; e iii) garantia de impugnação recursal.

Por fim, alude-se à aplicabilidade à criação e uso de robôs dos princípios da "beneficência", da "não-maleficência", do respeito pela "autonomia humana" e pela "justiça", além daqueles consagrados no art. 2.º do Tratado da União Europeia[453], nomeadamente a "não discriminação" e a "tolerância".

A Carta entrou em vigor em julho de 2021. Quase um ano depois, em junho de 2022, a Provedora de Justiça apresentou pedido de declaração de inconstitucionalidade[454] de parte do seu art.º 6.º, que determina a implementação de "Plano de Acção contra a

[452] Eis o teor do preceito: "Artigo 9.º Uso da inteligência artificial e de robôs. 1 - A utilização da inteligência artificial deve ser orientada pelo respeito dos direitos fundamentais, garantindo um justo equilíbrio entre os princípios da explicabilidade, da segurança, da transparência e da responsabilidade, que atenda às circunstâncias de cada caso concreto e estabeleça processos destinados a evitar quaisquer preconceitos e formas de discriminação. 2 - As decisões com impacto significativo na esfera dos destinatários que sejam tomadas mediante o uso de algoritmos devem ser comunicadas aos interessados, sendo suscetíveis de recurso e auditáveis, nos termos previstos na lei. 3 - São aplicáveis à criação e ao uso de robôs os princípios da beneficência, da não-maleficência, do respeito pela autonomia humana e pela justiça, bem como os princípios e valores consagrados no artigo 2.º do Tratado da União Europeia, designadamente a não discriminação e a tolerância". Fonte: https://dre.pt/dre/legislacao-consolidada/lei/2021-164870244 [Consult. 15 dez. 2022]

[453] O dispositivo assim estabelece, como se recorda: "Artigo 2.º A União funda-se nos valores do respeito pela dignidade humana, da liberdade, da democracia, da igualdade, do Estado de direito e do respeito pelos direitos do Homem, incluindo os direitos das pessoas pertencentes a minorias. Estes valores são comuns aos Estados-Membros, numa sociedade caracterizada pelo pluralismo, a não discriminação, a tolerância, a justiça, a solidariedade e a igualdade entre homens e mulheres". Fonte: https://www.europarl.europa.eu/about-parliament/pt/in-the-past/the-parliament-and-the-treaties [Consult. 15 dez. 2022].

[454] O teor do requerimento encontra-se em: https://www.provedor-jus.pt/provedora-de-justica-requer-fiscalizacao-da-constitucionalidade-de-normas-da-carta-portuguesa-de-direitos-humanos-na-era-digital/ [Consult. 15 dez. 2022].

Desinformação"[455] e também foi objeto de projeto de revogação apresentado pelo Partido Socialista[456].

3.2.5. Brasil

No Brasil, ainda não há qualquer marco legal específico a abordar a questão da IA de forma mais abrangente.

O que existem são algumas iniciativas legislativas ainda em fase de "gestação" no Congresso Nacional.

Dentre os projetos legislativos brasileiros, aquele que merece destaque pela sua abrangência e pelo seu avançado estágio de tramitação é o Projeto de Lei 21/2020.

O PL 21/20 foi apresentado no Plenário da Câmara dos Deputados brasileira pelo Deputado EDUARDO BISMARK em 03/02/2020 e pretende "criar o marco legal do desenvolvimento e uso da Inteligência Artificial (IA) pelo poder público, por empresas, entidades diversas e pessoas físicas".

[455] De acordo com LUIS MENEZES DE LEITÃO, "A existência de um plano de acção contra a desinformação pode ser justificada a nível da União Europeia, mas não nos parece que da mesma resulte a necessidade de uma norma como a do art. 6º da Lei 27/2021. Na verdade, esta disposição assume-se mais como uma restrição ao direito à liberdade de expressão do que um novo direito dos cidadãos, sendo certo que essa restrição surge com base em conceitos extremamente vagos como o de desinformação. Tal foi reconhecido pelo Presidente da República, no seu pedido de fiscalização sucessiva, ao estabelecer que 'o disposto no artigo 6º, ao procurar definir o conceito de desinformação e ao estabelecer mecanismos para a sua eliminação, poderia restringir o conteúdo do direito à liberdade de expressão, previsto no artigo 37º da Constituição'. Para além disso, o Presidente da República salienta que "as normas em causa, em especial as contidas nos n.ºs 1 a 4 do artigo 6º, conteriam um conjunto de conceitos vagos e indeterminados, de que são mero exemplo os seguintes: 'narrativa comprovadamente falsa ou enganadora"; "ameaça aos processos políticos democráticos, aos processos de elaboração de políticas públicas'; ou 'utilização de textos ou vídeos manipulados ou fabricados, bem como as práticas para inundar as caixas de correio eletrónico e o uso de redes de seguidores fictícios'. Da mesma forma, o Presidente da República salientou que também o n.º 6 do artigo 6º, ao prever que 'o Estado apoia a criação de estruturas de verificação de factos por órgãos de comunicação social devidamente registados e incentiva a atribuição de selos de qualidade por entidades fidedignas dotadas do estatuto de utilidade pública', poderia incorrer em inconstitucionalidade na medida em que, assentando nos conceitos indeterminados já referidos, previsse a atuação do Estado na criação de estruturas de verificação de factos cujo âmbito de atuação é desconhecido – e não o deveria ser no plano de uma lei restritiva – e cuja natureza ficaria também por esclarecer'.
Agora a Provedora de Justiça veio acrescentar a esse pedido do Presidente da República o da averiguação da inconstitucionalidade do nº5 desse artigo 6º, por considerar 'constitucionalmente inadmissível que alguém possa ser alvo de um processo de contraordenação por se limitar a exprimir ou difundir uma ideia, um pensamento ou mesmo determinado conteúdo informativo no ambiente digital". Em relação ao nº6 do art. 6º, a Provedora de Justiça considera não haver especificação da natureza e modo de apoio do Estado às estruturas de verificação de factos, ao contrário do que exige a Constituição em relação às leis restritivas de direitos fundamentais". Fonte: https://portal.oa.pt/comunicacao/imprensa/2022/06/14/a-inconstitucionalidade-da-carta-portuguesa-de-direitos-humanos-na-era-digital/ [Consult. 15 dez. 2022].

[456] Fonte: https://eco.sapo.pt/2022/06/21/ps-recua-e-propoe-revogar-artigo-6-o-da-carta-de-direitos-humanos-na-era-digital/ [Consult. 15 dez. 2022].

Foi aprovado em sessão deliberativa extraordinária em 29/09/2021 e seguiu para o Senado Federal, onde aguarda aprovação. Atualmente, tramita em conjunto com os Projetos de Lei 5.051, de 2019, e 872, de 2021, por tratarem de temas correlatos.

Em seus arts. 2º[457] e 9º[458], o projeto define inteligência artificial, de modo a demarcar o seu escopo normativo. No dispositivo seguinte, são elencadas aquelas que seriam as finalidades da IA no país, entre as quais podem ser referidas o desenvolvimento científico, tecnológico e econômico, assim como o aumento da competividade e a melhoria na prestação de serviços públicos[459].

Os fundamentos para o desenvolvimento e aplicação da IA no Brasil encontram-se delineados no art. 4º[460], que enaltece, dentre outros: a livre iniciativa e a livre concorrência; o respeito à ética, aos direitos humanos e aos valores democráticos; a livre manifestação de pensamento e a livre expressão da atividade intelectual, artística, científica e de comunicação;

[457] PL 21/20 da Câmara dos Deputados do Brasil - Art. 2º Para os fins desta Lei, considera-se sistema de inteligência artificial o sistema baseado em processo computacional que, a partir de um conjunto de objetivos definidos por humanos, pode, por meio do processamento de dados e de informações, aprender a perceber e a interpretar o ambiente externo, bem como a interagir com ele, fazendo predições, recomendações, classificações ou decisões, e que utiliza, sem a elas se limitar, técnicas como: I – sistemas de aprendizagem de máquina (machine learning), incluída aprendizagem supervisionada, não supervisionada e por reforço; II – sistemas baseados em conhecimento ou em lógica; III – abordagens estatísticas, inferência bayesiana, métodos de pesquisa e de otimização. Parágrafo único. Esta Lei não se aplica aos processos de automação exclusivamente orientados por parâmetros predefinidos de programação que não incluam a capacidade do sistema de aprender a perceber e a interpretar o ambiente externo, bem como a interagir com ele, a partir das ações e das informações recebidas.

[458] PL 21/20 da Câmara dos Deputados do Brasil - Art. 9º Para os fins desta Lei, sistemas de inteligência artificial são representações tecnológicas oriundas do campo da informática e da ciência da computação, competindo privativamente à União legislar e normatizar a matéria para a promoção de uniformidade legal em todo o território nacional, na forma do disposto no inciso IV do caput do art. 22 da Constituição Federal.

[459] PL 21/20 da Câmara dos Deputados do Brasil - Art. 3º A aplicação de inteligência artificial no Brasil tem por objetivo o desenvolvimento científico e tecnológico, bem como: I – a promoção do desenvolvimento econômico sustentável e inclusivo e do bem-estar da sociedade; II – o aumento da competitividade e da produtividade brasileira; III – a inserção competitiva do Brasil nas cadeias globais de valor; IV – a melhoria na prestação de serviços públicos e na implementação de políticas públicas; V – a promoção da pesquisa e desenvolvimento com a finalidade de estimular a inovação nos setores produtivos; e VI – a proteção e a preservação do meio ambiente.

[460] PL 21/20 da Câmara dos Deputados do Brasil - Art. 4º O desenvolvimento e a aplicação da inteligência artificial no Brasil têm como fundamentos: I – o desenvolvimento científico e tecnológico e a inovação; II – a livre iniciativa e a livre concorrência; III – o respeito à ética, aos direitos humanos e aos valores democráticos; IV – a livre manifestação de pensamento e a livre expressão da atividade intelectual, artística, científica e de comunicação; V – a não discriminação, a pluralidade, o respeito às diversidades regionais, a inclusão e o respeito aos direitos e garantias fundamentais do cidadão; VI – o reconhecimento de sua natureza digital, transversal e dinâmica; VII – o estímulo à autorregulação, mediante adoção de códigos de conduta e de guias de boas práticas, observados os princípios previstos no art. 5º desta Lei, e as boas práticas globais; VIII – a segurança, a privacidade e a proteção de dados pessoais; IX – a segurança da informação; X – o acesso à informação; XI – a defesa nacional, a segurança do Estado e a soberania nacional; XII – a liberdade dos modelos de negócios, desde que não conflite com as disposições estabelecidas nesta Lei; XIII – a preservação da estabilidade, da segurança, da resiliência e da funcionalidade dos sistemas de inteligência artificial, por meio de medidas técnicas compatíveis com os padrões internacionais e de estímulo ao uso de boas práticas; XIV – a proteção da livre concorrência e contra práticas abusivas de mercado, na forma da Lei nº 12.529, de 30 de novembro de 2011; e XV – a harmonização com as Leis nºs 13.709, de 14 de agosto de 2018 (Lei Geral de Proteção de Dados Pessoais), 12.965, de 23 de abril de 2014, 12.529, de 30 de novembro de 2011, 8.078, de 11 de setembro de 1990 (Código de Defesa do Consumidor), e 12.527 de 18 de novembro de 2011. Parágrafo único. Os códigos de conduta e os guias de boas práticas previstos no inciso VII do caput deste artigo poderão servir como elementos indicativos de conformidade.

a não discriminação, a pluralidade, o respeito às diversidades regionais, a inclusão e o respeito aos direitos e garantias fundamentais do cidadão; o estímulo à autorregulação; a segurança, a privacidade e a proteção de dados pessoais; o acesso à informação; a defesa nacional, a segurança do Estado e a soberania nacional; a preservação da estabilidade, da segurança, da resiliência e da funcionalidade dos sistemas de inteligência artificial, por meio de medidas técnicas compatíveis com os padrões internacionais e de estímulo ao uso de boas práticas; a proteção da livre concorrência e contra práticas abusivas de mercado.

O epicentro normativo do projeto talvez possa ser identificado em seu art. 5º[461], no qual estão relacionados os princípios que deveriam reger o uso da IA no Brasil. Inspirado em outros instrumentos (aqui já mencionados), o PL 21/20 alude, ilustrativamente, à finalidade benéfica a ser perseguida pela IA, sua centralidade no ser humano, além de sua neutralidade, transparência e segurança, como parâmetros a serem observados no desenvolvimento e implantação de soluções tecnológicas dotadas de inteligência artificial.

Esse corpo principiológico atuará com importante balizamento para os agentes públicos e privados, mas a ausência de qualquer alusão a mecanismos de coerção diretos ou indiretos fragiliza a sua eficácia normativa concreta.

[461] PL 21/20 da Câmara dos Deputados do Brasil - Art. 5º São princípios para o desenvolvimento e a aplicação da inteligência artificial no Brasil: I – finalidade benéfica: busca de resultados benéficos para a humanidade pelos sistemas de inteligência artificial; II – centralidade do ser humano: respeito à dignidade humana, à privacidade, à proteção de dados pessoais e aos direitos fundamentais, quando o sistema tratar de questões relacionadas ao ser humano; III – não discriminação: mitigação da possibilidade de uso dos sistemas para fins discriminatórios, ilícitos ou abusivos; IV – busca pela neutralidade: recomendação de que os agentes atuantes na cadeia de desenvolvimento e de operação de sistemas de inteligência artificial busquem identificar e mitigar vieses contrários ao disposto na legislação vigente; V – transparência: direito das pessoas de serem informadas de maneira clara, acessível e precisa sobre a utilização das soluções de inteligência artificial, salvo disposição legal em sentido contrário e observados os segredos comercial e industrial, nas seguintes hipóteses: a) sobre o fato de estarem se comunicando diretamente com sistemas de inteligência artificial, tal como por meio de robôs de conversação para atendimento personalizado on-line (chatbot), quando estiverem utilizando esses sistemas; b) sobre a identidade da pessoa natural, quando ela operar o sistema de maneira autônoma e individual, ou da pessoa jurídica responsável pela operação dos sistemas de inteligência artificial; c) sobre critérios gerais que orientam o funcionamento do sistema de inteligência artificial, assegurados os segredos comercial e industrial, quando houver potencial de risco relevante para os direitos fundamentais; VI – segurança e prevenção: utilização de medidas técnicas, organizacionais e administrativas, considerando o uso de meios razoáveis e disponíveis na ocasião, compatíveis com as melhores práticas, os padrões internacionais e a viabilidade econômica, direcionadas a permitir o gerenciamento e a mitigação de riscos oriundos da operação de sistemas de inteligência artificial durante todo o seu ciclo de vida e o seu contínuo funcionamento; VII – inovação responsável: garantia de adoção do disposto nesta Lei, pelos agentes que atuam na cadeia de desenvolvimento e operação de sistemas de inteligência artificial que estejam em uso, documentando seu processo interno de gestão e responsabilizando-se, nos limites de sua respectiva participação, do contexto e das tecnologias disponíveis, pelos resultados do funcionamento desses sistemas; VIII – disponibilidade de dados: não violação do direito de autor pelo uso de dados, de banco de dados e de textos por ele protegidos, para fins de treinamento de sistemas de inteligência artificial, desde que não seja impactada a exploração normal da obra por seu titular.

De outro lado, em seus arts. 6º[462] 7º[463] e 8º[464] estão tratadas as questões relativas ao emprego da IA e seus impactos na Administração Pública. Aqui merece realce o dever

[462] PL 21/20 da Câmara dos Deputados do Brasil - Art. 7º Constituem diretrizes para a atuação da União, dos Estados, do Distrito Federal e dos Municípios em relação ao uso e ao fomento dos sistemas de inteligência artificial no Brasil: I – promoção da confiança nas tecnologias de inteligência artificial, com disseminação de informações e de conhecimento sobre seus usos éticos e responsáveis; II – incentivo a investimentos em pesquisa e desenvolvimento de inteligência artificial; III – promoção da interoperabilidade tecnológica dos sistemas de inteligência artificial utilizados pelo poder público, de modo a permitir o intercâmbio de informações e a celeridade de procedimentos; IV – incentivo ao desenvolvimento e à adoção de sistemas de inteligência artificial nos setores público e privado; V – estímulo à capacitação e à preparação das pessoas para a reestruturação do mercado de trabalho; VI – estímulo a práticas pedagógicas inovadoras, com visão multidisciplinar, e ênfase da importância de ressignificação dos processos de formação de professores para lidar com os desafios decorrentes da inserção da inteligência artificial como ferramenta pedagógica em sala de aula; VII – estímulo à adoção de instrumentos regulatórios que promovam a inovação, como ambientes regulatórios experimentais (sandboxes regulatórios), análises de impacto regulatório e autorregulações setoriais; VIII – estímulo à criação de mecanismos de governança transparente e colaborativa, com a participação de representantes do poder público, do setor empresarial, da sociedade civil e da comunidade científica; e IX – promoção da cooperação internacional, mediante estímulo ao compartilhamento do conhecimento sobre sistemas de inteligência artificial e à negociação de tratados, acordos e padrões técnicos globais que facilitem a interoperabilidade entre os sistemas e a harmonização da legislação a esse respeito. Parágrafo único. Para fins deste artigo, o poder público federal promoverá a gestão estratégica e as orientações quanto ao uso transparente e ético de sistemas de inteligência artificial no setor público, conforme as políticas públicas estratégicas para o setor.

[463] PL 21/20 da Câmara dos Deputados do Brasil - Art. 6º Ao disciplinar a aplicação de inteligência artificial, o poder público deverá observar as seguintes diretrizes: I – intervenção subsidiária: regras específicas deverão ser desenvolvidas para os usos de sistemas de inteligência artificial apenas quando absolutamente necessárias para a garantia do atendimento ao disposto na legislação vigente; II – atuação setorial: a atuação do poder público deverá ocorrer pelo órgão ou entidade competente, considerados o contexto e o arcabouço regulatório específicos de cada setor; III – gestão baseada em risco: o desenvolvimento e o uso dos sistemas de inteligência artificial deverão considerar os riscos concretos, e as definições sobre a necessidade de regulação dos sistemas de inteligência artificial e sobre o respectivo grau de intervenção deverão ser sempre proporcionais aos riscos concretos oferecidos por cada sistema e à probabilidade de ocorrência desses riscos, avaliados sempre em comparação com:
a) os potenciais benefícios sociais e econômicos oferecidos pelo sistema de inteligência artificial; e b) os riscos apresentados por sistemas similares que não envolvam inteligência artificial, nos termos do inciso V deste caput;
IV – participação social e interdisciplinar: a adoção de normas que impactem o desenvolvimento e a operação de sistemas de inteligência artificial será baseada em evidências e precedida de consulta pública, realizada preferencialmente pela internet e com ampla divulgação prévia, de modo a possibilitar a participação de todos os interessados e as diversas especialidades envolvidas; V – análise de impacto regulatório: a adoção de normas que impactem o desenvolvimento e a operação de sistemas de inteligência artificial será precedida de análise de impacto regulatório, nos termos do Decreto nº 10.411, de 30 de junho de 2020, e da Lei nº 13.874, de 20 de setembro de 2019; e VI – responsabilidade: as normas sobre responsabilidade dos agentes que atuam na cadeia de desenvolvimento e operação de sistemas de inteligência artificial deverão, salvo disposição legal em contrário, pautar-se na responsabilidade subjetiva e levar em consideração a efetiva participação desses agentes, os danos específicos que se deseja evitar ou remediar e a forma como esses agentes podem demonstrar adequação às normas aplicáveis, por meio de esforços razoáveis compatíveis com os padrões internacionais e as melhores práticas de mercado. § 1º Na gestão com base em risco a que se refere o inciso III do caput deste artigo, a administração pública, nos casos de baixo risco, deverá incentivar a inovação responsável com a utilização de técnicas regulatórias flexíveis. § 2º Na gestão com base em risco a que se refere o inciso III do caput deste artigo, a administração pública, nos casos concretos em que se constatar alto risco, poderá, no âmbito da sua competência, requerer informações sobre as medidas de segurança e prevenção enumeradas no inciso VI do caput do art. 5º desta Lei, e respectivas salvaguardas, nos termos e nos limites de transparência estabelecidos por esta Lei, observados os segredos comercial e industrial. § 3º Quando a utilização do sistema de inteligência artificial envolver relações de consumo, o agente responderá independentemente de culpa pela reparação dos danos causados aos consumidores, no limite de sua participação efetiva no evento danoso, observada a Lei nº 8.078 de 11 de setembro de 1990 (Código de Defesa do Consumidor). § 4º As pessoas jurídicas de direito público e as de direito privado prestadoras de serviços públicos responderão pelos danos que seus agentes, nessa qualidade, causarem a terceiros, assegurado o direito de regresso contra o responsável nos casos de dolo ou culpa.

[464] PL 21/20 da Câmara dos Deputados do Brasil - Art. 8º As diretrizes de que tratam os arts. 6º e 7º desta Lei serão aplicadas conforme regulamentação do Poder Executivo federal por órgãos e entidades setoriais com competência

atribuído ao Poder Executivo Federal de monitorar a gestão do risco dos sistemas de inteligência artificial, para avaliar os riscos da aplicação e respectivas medidas de mitigação, além de estabelecer direitos, deveres e responsabilidades.

Por fim, há de ser ressaltado o explícito reconhecimento da importância da autorregulação, que vem mencionada em, pelo menos, dois preceitos (arts. 4º, VII, e 8º, III).

técnica na matéria, os quais deverão: I – monitorar a gestão do risco dos sistemas de inteligência artificial, no caso concreto, avaliando os riscos da aplicação e as medidas de mitigação em sua área de competência; II – estabelecer direitos, deveres e responsabilidades; e III – reconhecer instituições de autorregulação.

4. "Human rights by design"

4.1. Descrição

"*Human rights by design*" refere-se a uma via tecnoregulatória que consiste na introdução de definições apriorísticas a balizarem e restringirem o desenho de produtos ou artefatos tecnológicos com o objetivo de incorporar o respeito aos direitos humanos como traço arquitetural[465].

No concernente à IA, consistiria na incorporação de parâmetros estruturais alinhados aos direitos humanos na sua própria arquitetura algorítmica, de modo de obstar a violação dessas garantias mediante "travas" inseridas ao seu próprio "design" dos sistemas.

Outras expressões comumente associadas a tal abordagem seriam "*human rights centered design*", "*ethical-by-design*". "*design for values*"[466], "*embdeddedness of values in technology*", "*value-ladeness of technology*" e "*embodying values in technology*"[467], entre outras.

O termo "Design para Valores" talvez possa ser vista como um "gênero" de que seriam espécies os demais. Refere-se à construção de requisitos de design a partir de valores morais e sociais[468].

Abrangeria uma série de metodologias, como "*value sensitive design*"[469], "*values in design*"[470] e "*participatory design*" [471].

---

[465] AIZENBERG Evgeni & HOVEN Jeroen van den – *Designing for human rights in AI*, p. 2.
[466] AIZENBERG Evgeni & HOVEN Jeroen van den – *Op. cit.*, p. 2.
[467] FLANAGAN, Mary & HOWE, Daniel C. & NISSENBAUM, Helen – *Embodying Values in Technology – Theory and Practice*, p. 322.
[468] "*The term Design for Values (Van den Hoven et al., 2015) refers to the explicit translation of moral and social values into context-dependent design requirements. It encompasses under one umbrella term a number of pioneering methodologies, such as Value Sensitive Design (Friedman, 1997; Friedman and Hendry, 2019; Friedman et al., 2002), Values in Design (Flanagan et al., 2008; Nissenbaum, 2005) and Participatory Design (Schuler and Namioka, 1993)*"AIZENBERG Evgeni & HOVEN Jeroen van den – *Op. cit.*, p. 2.
[469] "*We shall see that value sensitive design is a formative framework (Rasmussen 1997; Vicente 1999), a framework that bounds and guides design processes. It frames and nudges, prompting designers to attend to value-oriented aspects of design situations and to pursue certain kinds of work. In short, value sensitive design places a distinctive shaping influence on designers. Formative frameworks differ in purpose from prescriptive ones, which prescribe that certain steps be followed to frame a design problem and to arrive at a solution*." HENDRY, David G. & FRIEDMAN, Batya & BALLARD, Stephanie – *Value sensitive design as a formative framework*, p. 1.
[470] "*The idea that values may be embodied in technical systems and devices (artifacts)*" NISSENBAUM, Helen – *Values in Technical Design*.
[471] "*Participatory design is a democratic process for design (social and technological) of systems involving human work, based on the argument that users should be involved in designs they will be using, and that all stakeholders, including and especially users, have equal input into interaction design* (Muller & Kuhn, 1993)" HARTSON, Rex & PYLA, Pardha – *History and Origins of Participatory Design*.

Não se trata de uma ideia propriamente nova, como meio para contenção de riscos jurídicos decorrentes de ferramentas computacionais[472].

Cuida-se de um importante contributo para o desenho de qualquer artefato tecnológico com potencial de produzir impactos socialmente relevantes, de forma a incorporar não apenas valores de natureza instrumental ou funcional (v.g. eficiência, segurança, confiabilidade e "usabilidade"), mas também os de caráter substancial (v.g. relativos aos aspectos social, moral, político e jurídico)[473].

Contudo, a ideia segundo a qual valores (interesses ou direitos) dessa segunda ordem possam estar incorporados em produtos, processo ou sistemas não reúne consenso entre o que transitam nas áreas de interseção entre ciência, tecnologia e ética[474].

4.2. Possíveis soluções metológicas

Dadas as dificuldades em se parametrizar com alguma objetividade os valores e interesses[475] a serem internalizados pelos agentes artificialmente inteligentes, tem sido cogitadas soluções que envolvem, ilustrativamente, um fluxo de especificação de valores, de um lado, e uma espécie de aprendizado de máquinas por inferência, de outro[476].

Apenas para que se possa aquilatar as possibilidades técnicas de implementação, podem ser referidas algumas estratégias destinadas a garantir o alinhamento entre valores e direitos humanos e metas perseguidas por sistemas de IA, além de conter riscos de vulnerações mais sensíveis, sobretudo por sistemas mais avançados.

4.1.1. "Value Specification"

A gradual "tradução" de valores abstratos em requisitos de design é conhecida como "*value specification*", i.e., especificação de valores, que pode ser visualmente mapeada de forma a produzir uma hierarquia axiológica contextual[477].

---

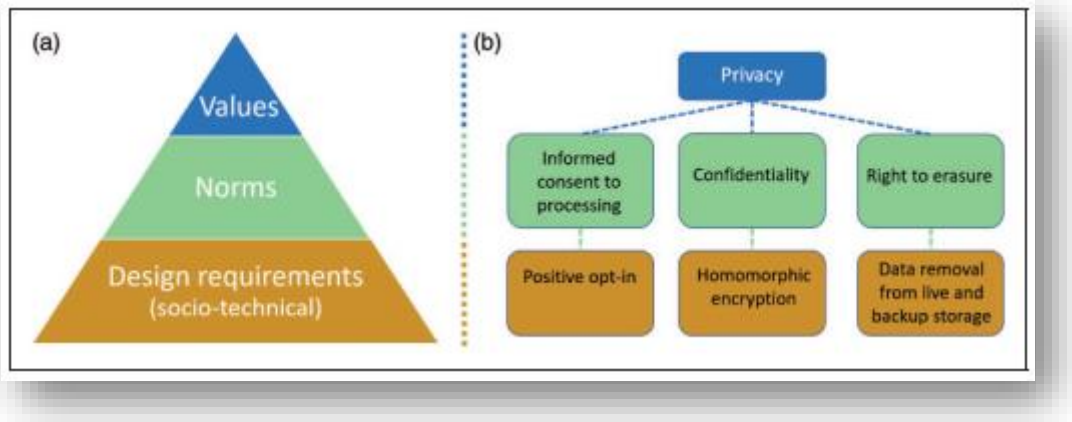
[472] Entre outras referências, pode ser mencionada a obra de FRIEDMAN publicada na década de 1990: FRIEDMAN, Batya – *Human Values and the Design of Computer Technology* (1997).
[473] FLANAGAN, Mary & HOWE, Daniel C. & NISSENBAUM, Helen – *Op. cit.*, p. 322.
[474] NISSENBAUM, Helen – *Values in Technical Design*, p. 66.
[475] Eis o que afirma BOSTROM, ilustrativamente: "*It seems completely impossible to write down a list of everything we care about*". FITCH, Andy – *Letter from Utopia: Talking to Nick Bostrom*.
[476] CHRISTIAN, Brian – *The Alignment Problem*, p. 247
[477] AIZENBERG Evgeni & HOVEN Jeroen van den - *Designing for human rights in AI*, p. 3.

Haveria, essencialmente, três etapas: i) definição dos valores a serem considerados; ii) elaboração de normas a partir dos valores definidos – trata-se da indicação das propriedades e capacidades de que o sistema deveria ser dotado para concretizar os valores escolhidos; e iii) descrição dos requisitos de design socio-tecnológicos – tais exigências seriam, ao mesmo tempo, sociais e técnicas, por refletirem a interpelação entre o homem e seus artefatos tecnológicos[478].

Em uma representação visual, pode-se assim descrever o processo e seus resultados:

Figura 7.

Representação visual da "value specification" [479]



No caso do valor privacidade, ilustrativamente, seriam definidas as seguintes normas: necessidade de informar consentimento, confidencialidade e direito de apagar os dados[480]. Já os requisitos operacionais de design consistiriam, exemplificativamente, em: pemissão do usuário para o processamento dos dados pessoais ("positive opt-in"), encriptação dos dados[481] e remoção dos dados dos bancos de dado e backups[482].

Para alcançar tais definições, poder-se-ia observar uma metodologia tripartite, a partir de três modalidades de investigação. não estanques nem lineares: i) conceitual – identificação

[478] AIZENBERG Evgeni & HOVEN Jeroen van den – *Op. cit.*, p. 3.
[479] *Id., ibid.*
[480] *Id., ibid.*
[481] A citada encriptação homomórfica consiste, fundamentalmente, em técnica avançada de codificação dos dados, de modo a preservar sua confidencialidade, sem impedir o seu manuseio pelos sistemas. Em linguagem mais técnica: "*Homomorphic encryption is the conversion of data into ciphertext that can be analyzed and worked with as if it were still in its original form. Homomorphic encryption enables complex mathematical operations to be performed on encrypted data without compromising the encryption*". Fonte: https://www.techtarget.com/searchsecurity/definition/homomorphic-encryption [Consult. 3 nov. 2022]
[482] AIZENBERG Evgeni & HOVEN Jeroen van den – *Op. cit.*, p. 3.

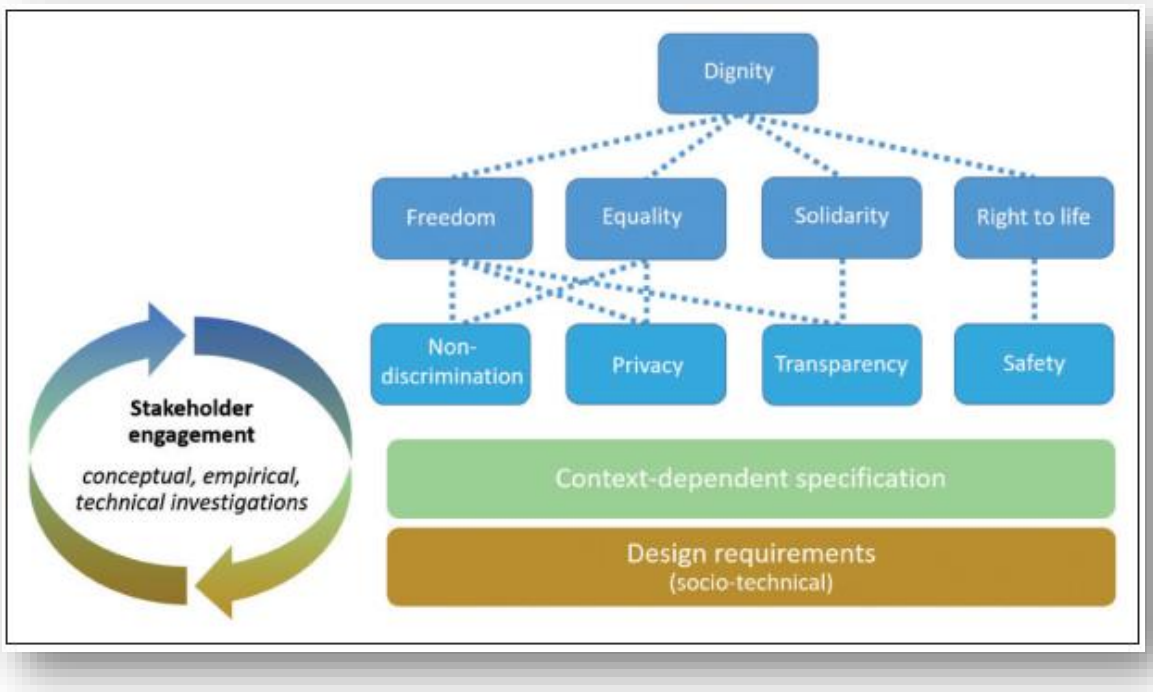
dos "stakeholders" e dos valores impactados pelo artefato tecnológico; ii) empírica – análise das necessidades, visões e experiências de todos os envolvidos quanto aos impactos da tecnologia e suas implicações; e iii) técnica – avaliação das soluções técnicas para implementar as normas concebidas a partir dos valores definidos[483].

4.1.1.1. Aplicação exemplificativa no cenário europeu

No cenário europeu, a referida trilha metodológica poderia considerar, v.g., os direitos humanos que figuram com pilares centrais da Carta Europeia de Direitos Humanos: vida, dignidade, liberdade, igualdade e solidariedade (arts.º 1.º, 2.º, 6.º, 8.º, 11.º, 21.º e 34.º)[484]:

Figura 8.

Aplicação no cenário europeu da "value specification" [485]



A partir desse processo, seria possível conceber formas de certificação para restringir o acesso ao mercado aos sistemas cuja arquitetura fosse condizente com os mencionados

[483] AIZENBERG Evgeni & HOVEN Jeroen van den – *Op. cit.*, p. 4.
[484] AIZENBERG Evgeni & HOVEN Jeroen van den – *Op. cit.*, p. 6. As linhas pontilhadas sugerem as interelações entre os aludidos direitos.
[485] AIZENBERG Evgeni & HOVEN Jeroen van den – *Op. cit.*, p. 6. As linhas pontilhadas sugerem as interelações entre os aludidos direitos.

valores, ao cumprir os requisitos normativos impostos, tal como se exige de outros tecnologias sujeitas a padrões e "standards" qualitativos[486].

Seria necessário, nessa perspectiva, estabelecer uma espécie de estudo de impacto[487] e uma "*due dilligence*" a respeito dos possíveis riscos aos direitos humanos resultantes da implementação de sistemas de IA em desenvolvimento, antes de poderem ser efetivamente utilizados[488].

A matriz de riscos definida na proposta europeia de regulação do tema (analisada no terceiro capítulo da presente dissertação) talvez possa guiar semelhante abordagem, ao indicar os níveis de "tolerância" às vulnerabilidades contidas em soluções de IA.

4.1.2. "*Coherent Extrapolated Volition*"

"*Coherent Extrapolated Volition*" (CEV) ou, em tradução literal, "Volição Extrapolada Coerente"[489] seria uma abordagem cuja ideia central consiste na tentativa de inferir da vontade humana elementos para definir o seu núcleo volitivo essencial, a partir de uma análise coerente de aspectos implícitos e contextuais de modo a conduzir o processo de tomada de decisão de agentes artificialmente inteligentes voltados para a concretização de tal vontade[490].

---

[486] AIZENBERG Evgeni & HOVEN Jeroen van den – *Op. cit.*, p. 10.

[487] "*The Algorithmic Impact Assessment (AIA) framework* […] *is designed to support affected communities and stakeholders as they seek to assess the claims made about these systems, and to determine where – or if – their use is acceptable*. […] *Algorithmic Impact Assessments draw directly from impact assessment frameworks in environmental protection, data protection, privacy, and human rights policy domains*". REISMAN Dillon & SCHULTZ, Jason & CRAWFORD, Kate, WHITTAKER, Meredith - *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability*. "*An AIA-based regulatory framework would require the creator of an algorithmic system to assess its potential socially harmful impacts before implementation and create documentation that can be used later for accountability and future policy development*". SELBST, Andrew D. - *An Institutional View of Algorithmic Impact Assessments*, p. 117.

[488] Id., ibid.

[489] "*In poetic terms, our coherent extrapolated volition is our wish if we knew more, thought faster, were more the people we wished we were, had grown up farther together; where the extrapolation converges rather than diverges, where our wishes cohere rather than interfere; extrapolated as we wish that extrapolated, interpreted as we wish that interpreted*". **YUDKOWSKY, Eliezer –** *Coherent Extrapolated Volition*.

[490] "*A very simple example of extrapolated volition might be to consider somebody who asks you to bring them orange juice from the refrigerator. You open the refrigerator and see no orange juice, but there's lemonade. You imagine that your friend would want you to bring them lemonade if they knew everything you knew about the refrigerator, so you bring them lemonade instead. On an abstract level, we can say that you 'extrapolated' your friend's 'volition', in other words, you took your model of their mind and decision process, or your model of their 'volition', and you imagined a counterfactual version of their mind that had better information about the contents of your refrigerator, thereby 'extrapolating' this volition*". Fonte: https://arbital.com/p/cev/ [Consult. 14 out. 2022].

Cuida-se de um modelo destinado sobretudo a sistemas mais avançados de IA (dotados de inteligência artificial geral ou superinteligência) e que partiria de três premissas ou condições para definir em que consistira a vontade do sujeito humano (SH) a ser "satisfeita" pelo agente artificial (AA)[491].

O AA deveria tentar identificar o conteúdo volitivo de SH a partir de um cenário suposto em que: i) SH dispusesse de todas as informações com que AA contasse; ii) SH pudesse raciocinar tão rapidamente quanto AA para ponderar todos os argumentos pertinentes; e iii) SH conhecesse plenamente a si mesmo e tivesse suficiente autodomínio e coerência[492].

Nesse contexto, em seu nível individual[493], a técnica da extrapolação volitiva consideraria três direções ou vetores: i) maior conhecimento – disponibilidade de conhecimento mais preciso de fatos declarativos e resultados esperados; ii) maior consideração de argumentos - capacidade de sopesamento de mais argumentos possíveis e avaliação de sua validade; e iii) maior "refletividade" - maior autoconhecimento e maior autocontrole[494].

Há imensas discussões acerca da factibilidade, do alcance, da delimitação semântica e de vários outros aspectos a respeito de tais caminhos metodológicos, cujo aprofundamento refoge aos limites da presente investigação.

Seja como for, a sua existência representa uma importante evidência dos esforços acadêmicos que tem sido envidados para incorporar o respeito a valores e direitos no "design" de sistemas de IA de maior sofisticação e capacidade cognitiva.

4.1.3. Controle de capacidades ("*Capability control*")

Ainda na seara dos sistemas de IA mais avançados (especialmente no nível da ainda hipotética "*Super AI*"), para implementação de mecanismos tecnoregulatórios tem-se cogitado também da introdução de algumas restrições de capacidade.

---

[491] Fonte: https://arbital.com/p/cev/ [Consult. 14 out. 2022].
[492] Fonte: https://arbital.com/p/cev/ [Consult. 14 out. 2022].
[493] "*Extrapolated volition is the metaethical theory that when we ask 'What is right?', then insofar as we're asking something meaningful, we're asking 'What would a counterfactual idealized version of myself want* if it knew all the facts, had considered all the arguments, and had perfect self-knowledge and self-control?' (As a metaethical theory, this would make 'What is right?' a mixed logical and empirical question, a function over possible states of the world.)*" Fonte: https://arbital.com/p/cev/ [Consult. 14 out. 2022].
[494] Fonte: https://arbital.com/p/cev/ [Consult. 14 out. 2022].

Seria uma forma de limitar determinar aspectos relativos às possibilidades de intervenção e interação dos agentes artificialmente inteligentes com o "mundo exterior".

Entre os métodos concretamente concebidos, podem ser citados alguns conjuntos de soluções[495]:

i) encaixotamento ("*boxing*") – confinamento dos sistemas, de tal forma que apenas poderiam afetar o mundo externo por meio de algum canal restrito e pré-aprovado, a abranger a contenção sob a perspectiva informacional (tráfego de dados) e também física (confinamento dos equipamentos);

ii) incentivo[496] – construção de um modelo de incentivos[497] apropriados para evitar o desalinhamento de valores, que pode envolver a "integração social" em um mundo de entidades "igualmente poderosas", além do uso de "tokens de recompensa" (criptográficos) ou a "captura antrópica"[498];

iii) restrição de desenvolvimento ("*stunting*"): imposição de limitações nas capacidades cognitivas do sistema ou em sua capacidade de afetar os principais processos internos;

iv) armadilhas ("*tripwires*") – realização de testes de diagnóstico no sistema (sem sua detecção), associados a mecanismos de desativação, caso alguma ameaça seja identificada.

4.1.4. Seleção de motivações

Podem ser ainda lembradas técnicas relativas à definição de motivações especificamente destinadas à vinculação dos objetivos perseguidos por agentes de IA[499] e valores humanos centrais.

A "seleção de motivações" reúne diversas metodologias, entre as quais podem ser referidas[500]:

---

[495] Nick Bostrom - *Superintelligence*, Capitulo IX.

[496] No original: "*The system is placed within an environment that provides appropriate incentives. This could involve social integration into a world of similarly powerful entities. Another variation is the use of methods (cryptographic) reward tokens. "Anthropic capture" is also a very important possibility but one that involves esoteric considerations.*"

[497] Parâmetros que guiam os sistemas para determinadas direções, a partir de "estímulos" positivos e negativos, definidos matematicamente.

[498] A captura antrópica consistiria em método de controle de capacidade no qual a inteligência artificial avançada seria induzida a partir da premissa de estar em uma simulação e, como tal, tentaria se comportar de modo a ser recompensada por seus simuladores. Fonte: https://forum.effectivealtruism.org/topics/anthropic-capture/history [Consult. 30 nov. 2022]

[499] Parte das abordagens mencionadas concerne a sistemas mais avançados, ainda não desenvolvidos de forma plena, razão pela qual referem-se a preocupações de longo prazo.

[500] BOSTROM, Nick - *Superintelligence*, Capitulo IX.

i) especificação direta ("*direct specification*") – atribuição de motivações especificadas de forma direta, a partir de uma perspectiva consequencialista ou de um conjunto de regras próprias;

ii) domesticidade ("*domesticity*") – estabelecimento de motivações destinadas a limitar severamente o escopo de atuação e "ambições" do agente;

iii) normatividade indireta ("*indirect normativity*") – definição de um corpo de princípios baseados em regras ou consequencialistas, para orientar, indiretamente, as ações do sistema;

iv) aprimoramento ("*augmentation*") – identificação de sistemas já dotados de motivações substancialmente humanas ou benevolentes para aprimorar suas capacidades cognitivas e assim torná-lo superinteligente.

4.2. Possibilidades

4.2.1. Eficiência como estratégia de tecnoregulação

A introdução de travas sistémicas como forma de conter os riscos decorrentes das vulnerabilidades da IA quanto aos direitos humanos consiste em abordagem tecnoregulatória[501], no sentido já empregado no presente texto.

Cuida-se, pois, do uso da tecnologia como instrumento para regular o comportamento humano ou mesmo das "máquinas" inteligentes, as quais podem atuar com diferentes níveis de autonomia, como visto.

A eficiência, em tese, desse método regulatório depende, em grande medida, da capacidade de promover a inserção dos valores fundantes dos direitos humanos no processo de desenvolvimento e implantação dos sistemas de IA.

Uma vez assegurada a transposição fiel de tais valores, sem comprometimento de seu sentido original, as possibilidades de vulneração dos direitos humanos reduzir-se-iam sensivelmente. Isso porque haveria uma inviabilidade operacional inscrita no próprio "código-fonte" ou no algoritmo de tais sistemas a afastar de forma quase absoluta esse risco.

[501] Vale notar que a própria IA pode ser adotada como ferramenta para promover ou ampliar a tecnoregulação do comportamento humano, de forma geral. Entre outros: ELIOT, Lance B. – *Using AI As A Techno-Regulatory Enforcement Tool*; e VAN DEN BERG, Bibi – *Robots as Tools for Techno-Regulation*.

Portanto, como abordagem regulatória, "*human rights by design*" pode ser vista como uma via razoavelmente segura para evitar o "desalinhamento" dos sistemas de IA em face dos direitos humanos.

A definição dos parâmetros concretos a serem observados pelos desenvolvedores poderia ser objeto de instrumentos de "*hard law*" ou de "*soft law*", assim como de mecanismos de autoregulação regulada.

Há, ademais, um instrumento que tende a ser bastante útil para o teste de eficiência de tais mecanismos tecnoregulatórios, conhecido como "*regulatory sandbox*" e já antes mencionado na presente dissertação.

4.2.2. "*Sandboxes*" e "*Regulatory Sandboxes*" ("caixas de areia" regulatórias)

"*Sandboxes*" (caixas de areia) constituem um ambiente de testes seguros de programas de computador que representam algum risco relevante[502], um expediente utilizado há algum tempo por desenvolvedores privados e públicos.

Já as caixas de areia regulatória representam um conceito razoavelmente recente, que teria surgido em 2014 estimular novas tecnologias relativas ao mercado financeiro ("*fintech products*")[503].

Trata-se de uma forma de não frustrar a evolução tecnológica sobretudo por um excesso de rigor regulatório. Implementa-se por meio de um "experimento legislativo" ("*legal experiment*") no qual as regras nacionais são temporariamente flexibilizadas para promover a inovação em determinado seguimento[504].

O período experimental de abrandamento de incidência normativa seria objeto de monitoramento próximo por parte de uma autoridade competente, que faria as intervenções que reputasse necessárias em caso de alguma situação mais emergencial[505].

A Proposta Europeia de Regulação da IA contempla, expressamente, essa possibilidade[506].

---

[502] Fonte, entre outras: https://www.proofpoint.com/us/threat-reference/sandbox [Consult. 2 dez. 2022].
[503] POP, Florina – *Sandboxes for Responsible Artificial Intelligence*.
[504] RANCHORDAS, Sofia - *Experimental Regulations for AI: Sandboxes for Morals and Mores*, p. 12.
[505] POP, Florina – *Op. cit.*
[506] Em seu art.º 53.º assim prevê: "*1. AI regulatory sandboxes established by one or more Member States competent authorities or the European Data Protection Supervisor shall provide a controlled environment that facilitates the development, testing and validation of innovative AI systems for a limited time before their placement on the market or putting into service pursuant to a specific plan. This shall take place under the direct supervision and guidance by the competent authorities with a view to ensuring compliance with the requirements of this Regulation and, where relevant, other Union and Member States legislation supervised within the sandbox. 2. Member States shall ensure that to the extent the innovative AI systems involve the processing of personal data or otherwise fall*

Caso pudesse ser executada a estratégia regulatória aqui proposta (tecnoregulação via "*human rights by design*"), uma forma de verificar sua eficiência seria precisamente a de adotar a solução de "*regulatory sandboxes*".

Com isso, seriam selecionados determinados sistemas específicos para avaliar se, afastado todo o acervo normativo convencional, as travas introduzidas na sua arquitetura seriam suficientes para impedir a agressão aos direitos humanos.

Para alcançar semelhante desiderato, evidentemente, seria imperioso o acompanhamento por parte das autoridades competentes de supervisão, as quais mediriam a eficiência tecnoregulatória e promoveriam as eventuais intervenções para conter potenciais riscos de natureza emergencial.

4.3. Limites

4.3.1. Dificuldades epistemológicas

A conhecida separação ou mesmo distanciamento entre ciências "humanas" ou "sociais", de um lado, e "exatas" ou "tecnológicas", de outro, conduz à dificuldade de compreensão de aspectos conceituais e contextuais por parte tanto dos responsáveis pelo desenvolvimento técnico de sistemas de IA[507] quanto dos operadores do Direito.

Quanto aos profissionais de Tecnologia da Informação (TI), ao avaliar as diferentes opções arquiteturais ou de design, consideram, em regra, elemento de natureza estritamente técnica ou operacional, quanto à quais mantém desenvoltura, graças à sua formação específica[508].

Entretanto, ao se verem obrigado a introduzir elementos que traduzam ou concretizam valores de outra ordem (v.g. jurídica, moral, econômica ou social), encontram-se "desconfortáveis" ao serem compelidos a se valer de conhecimentos cujo alcance lhes escaparia[509].

---

under the supervisory remit of other national authorities or competent authorities providing or supporting access to data, the national data protection authorities and those other national authorities are associated to the operation of the AI regulatory sandbox. 3. The AI regulatory sandboxes shall not affect the supervisory and corrective powers of the competent authorities. Any significant risks to health and safety and fundamental rights identified during the development and testing of such systems shall result in immediate mitigation and, failing that, in the suspension of the development and testing process until such mitigation takes place."

[507] FLANAGAN, Mary & HOWE, Daniel C. & NISSENBAUM, Helen – *Op. cit.*, p. 324.

[508] *Id. Ibid*.

[509] *Id. Ibid*.

O mesmo pode ser afirmado, aliás, na direção oposta, i.e., quanto aos profissionais de Direito, os quais também se vêem diante de um terreno inóspito e árido ao realizarem a análise de aspectos estruturais dos sistemas de IA para compreenderem os eventuais impactos jurídicos de determinadas escolhas técnicas.

Seja como for, poderia haver um abismo epistemológico quase intransponível entre os dois âmbitos de avaliação. Sua superação supõe um ingente esforço para promover uma aproximação também ao nível acadêmico de áreas do conhecimento tradicionalmente afastadas e estanques.

Para conter os riscos mais iminentes, a criação de forças de trabalho multidisciplinares apresenta-se como uma das soluções para enfrentar os desafios de curto prazo, quanto aos sistemas já implementados ou em vias de implementação.

4.3.2. Dificuldades "implementacionais"

4.3.2.1. Consenso

O cenário ético hodierno em torno da IA encontra-se marcado por uma alta heterogeneidade quantos aos valores, princípios e direitos a serem resguardados[510].

De acordo com estudo conduzido no ano de 2021 em quase uma centena documentos (predominantemente instrumentos de "*soft law*") que tratam do design ético e da implantação de tecnologias de IA, seria possível mensurar o nível de divergência/convergência quanto a tais aspectos, a partir da menção a determinados princípios. Os dados colhidos a partir da análise de uma tríplice fonte (atores públicos, entes privados e especialistas) foram assim tabulados:

[510] RUDSCHIES, Catharina & SCHNEIDER, Ingrid & SIMON, Judith - *Value Pluralism in the AI Ethics Debate – Different Actors, Different Priorities*, p. 2. Para um outro levantamento panorâmico de princípios e valores: FJELD, Jessica *et ali – Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*.

Tabela 5.

Nível de divergência/convergência sobre valores e princípios em IA[511]

Principles	Proportion of ethics guidelines mentioning principle X - all three actor groups combined	Proportion of ethics guidelines mentioning principle X - public actors	Proportion of ethics guidelines mentioning principle X - expert actors	Proportion of ethics guidelines mentioning principle X - private actors
Principles of "common ground"				
transparency	95%	100%	92.9%	92.7%
privacy	87.5%	85.7%	92.9%	83.3%
non-discrimination and fairness	82.5%	92.9%	78.6%	75%
responsibility	77.5%	71.4%	92.9%	66.7%
explainability	77.5%	92.9%	85.7%	50%
safety and security	75%	78.6%	78.6%	66.7%
accountability	75%	78.6%	85.7%	58.3%
non-maleficence	67.5%	71.4%	85.7%	41.7%
Human and citizen rights	62.5%	78.6%	64.3%	41.7%
Principles without an overall "majority"				
freedom	47.5%	71.4%	64.3%	0%
dignity	32.5%	42.9%	50%	0%
sustainability	37.5%	57.1%	35.7%	16.7%
democracy and rule of law	42.5%	71.4%	42.9%	8.3%
contestability	40%	50%	57.1%	8.3%
accuracy and completeness	40%	50%	50%	16.7%

Os percentuais elevados apontam para a maior convergência, de modo a agrupar aqueles princípios com maior aderência na parte superior da tabela ("*principles of 'common ground*'") e os de menor consenso na parte inferior do quadro ("*principles without an overall 'majority*'").

Entre os primeiros, despontam, como se vê, a transparência, privacidade e não-discriminação (ou justiça/equidade – "*fairness*"). Já entre os últimos, podem ser identificados, entre outros, a liberdade, a dignidade e a democracia.

Impressiona a circunstância de que, apesar dos objetivos semelhantes de encontrar um "terreno comum", "temas abrangentes" ou "requisitos mínimos" para o design ético e a implantação de tecnologias de IA, as análises chegam a resultados divergentes[512].

[511] RUDSCHIES, Catharina & SCHNEIDER, Ingrid & SIMON, Judith – *Op. cit.*, p. 8.

[512] Em tradução livre: "Hagendorff chega à conclusão de que existem três princípios (privacidade, justiça e responsabilidade) que podem ser chamados de requisitos mínimos para uma IA ética, pois foram mencionados por pelo menos 90% das diretrizes que examinou. Floridi et ai. toma como base os princípios bioéticos da não maleficência, da beneficência, da autonomia e da justiça e os complementa com o princípio da explicabilidade,

Mesmo quanto a valores primários e direitos humanos, há divergências especialmente agudas nos tempos hodiernos, tanto no âmbito doméstico (nacional) quanto no externo (internacional)[513].

A "fratura" ideológica e polarização[514] percebida nos debates políticos e sociais em uma quantidade expressiva de países democráticos repercute no modo como são interpretados os alcances semânticos de garantias como a vida (v.g. aborto), liberdade de expressão (v.g. discurso do ódio nas redes sociais), autonomia (v.g. eutanásia), entre outros.

De outra parte, afigura-se como truísmo considerar as profundas dissonâncias entre Estados de tradições jurídicas heterogêneas no concernente ao modo como os direitos humanos são percebidos, interpretados e assegurados[515].

---

incluindo a inteligibilidade e a responsabilidade. Jobin et al. chamam transparência, privacidade, justiça e equidade, responsabilidade e não maleficência de "princípios éticos fundamentais" para a IA. E Fjeld et al. agrupar os princípios éticos sob os seguintes temas: transparência e explicabilidade, privacidade, não discriminação e justiça, segurança e proteção, responsabilidade, controle humano da tecnologia e promoção de valores humanos". Embora existam algumas sobreposições em todos esses conjuntos de princípios, a maior consonância apontaria para um "princípio fundamental" para a IA: justiça e equidade. RUDSCHIES, Catharina & SCHNEIDER, Ingrid & SIMON, Judith – *Op. cit.*, p. 2.

[513] No cenário internacional, uma das mais relevantes "caixas de ressonância" de tais discrepâncias é o Conselho de Direitos Humanos (CDH) da Organização das Nações Unidas (ONU). A esse propósito, eis a descrição de algumas mudanças mais recentemente ocorridas na forma como a questão da universalidade vem sendo desafiada perante o órgão: "*Discussions of human rights universality within the Human Rights Council mirror broader debates regarding the relationships between States, communities and individuals in light of changes in political, economic, social and cultural ideologies and institutions. Prior to the Council's establishment, discourses on the universality of human rights tended to focus on national or regional communitarian value systems that were thought to be antithetical to the individual rights guarantees contained in international human rights instruments. Since the Vienna World Conference on Human Rights in 1993, the focus of these discourses has shifted, and as reflected in the trends apparent within the Human Rights Council's debates, recent challenges to human rights universality have taken more subtle forms, including appeals (whether implicit or explicit) to exceptionalism on the basis of particular security threats or prevailing economic and environmental conditions. The question of cultural relativism is still present, however, and is most apparent in discussions concerning freedom of religion, women's sexual and reproductive rights, 'traditional values', the protection of the family and the rights of LGBTI persons*". GENEVA ACADEMY – *Universality in the Human Rights Council: Challenges and Achievements*, p. 6.

[514] Curiosamente, a IA tem servido de ferramenta tanto para gerar e amplificar os dissensos quanto para detetar e combater os mecanismos indutores de radicalização (v.g. "fake news"). Quanto à primeira possibilidade, entre outras fontes, podem ser citadas: MODGIL, Sachin & SINGH, Rohit Kumar & GUPTA, Shivam & DENNEHY, Denis – *A Confirmation Bias View on Social Media Induced Polarisation During Covid-19*; STEIN, Yaakov – *Artificial Intelligence in Social Networks Produces Polarization*; e CALICE, Mikhaila N. & BAO, Luye & SCHEUFELE, Dietram A. – *Polarized platforms? How partisanship shapes perceptions of "algorithmic news bias"*. Já quanto à segunda perspectiva: YANKOSKI, Michael & WENINGER, Tim & SCHEIRER, Walter - *An AI early warning system to monitor online disinformation, stop violence, and protect elections* e BUKHARI, S. M. A. H. & KHAN, M. U. S. & KHAN, T. & ALI, M. - *Social media, AI, and political affiliation*; LOREGGIA, Andrea & MATTEI, Nicholas & QUINTARELLI, Stefano – *Artificial Intelligence Research for Fighting Political Polarisation: A Research Agenda*; e YANKOSKI, Michael & THEISEN, William & VERDEJA, Ernesto & SCHEIRER, Walter J. – *Artificial Intelligence for Peace: An Early Warning System for Mass Violence*.

[515] Para uma amostra, entre tantas possíveis, de tais dissonâncias, vale a leitura de entrevista concedida acerca do tema da universalidade dos direitos humanos, por professor chinês da Unviersidade de Zhengzhou. Sua avaliação do suposto eurocentrismo revela-se especialmente ilustrativa: "[…] *Eurocentrism actually implies a way of thinking: the assumption that what is universal is something Western, and that non-Western things are particular.*

Embora seus instrumentos contemporâneos tenham "nascido" sob o signo da universalidade e do consenso entre os países que compunham as Nações Unidas, os direitos humanos ainda desafiam o consenso sobretudo em face das diversidades socioeconômicas, culturais e mesmo religiosas[516].

Portanto, entre as dificuldades qualificadas como "operacionais" (relativas à implementação das ferramentas tecnoregulatórias mencionadas), uma das mais relevantes seria aquela concernente à formação de convergência em torno do conjunto mínimo de direitos humanos a serem considerados, assim como do sentido concreto a ser atribuído a cada um deles[517].

A superação desse desafio passaria pelo aprofundamento do diálogo entre os Estados representativos de cada um dos grandes blocos socio-culturais e econômicos que formam a comunidade internacional.

Os meios específicos para alcançar semelhante desiderato, contudo, encontram-se fora do alcance da presente investigação[518].

Há de se referir que, no contexto europeu, a proposta de regulação da IA[519] prestigia, naturalmente, os direitos fundamentais[520] contidos na Carta Europeia, entre os quais se

---

Why is it that when talking about the universality and the particularity of human rights, people almost always refer to 'Asian values', 'Islam' or 'Confucianism' but rarely to the 'European or American way' as examples of particularity? This assumption essentially equates the West with universality. What's more, it exists not only in Western society as it has influenced the thinking of non-Western intellectuals who very often unconsciously fall in the Eurocentric way of thinking" Disponível em: https://www.graduateinstitute.ch/communications/news/universality-human-rights [Consult. 23 nov. 2022].

[516] Ilustrativamente, eis o que ponderam autores asiáticos em torno do tema: "*Among human rights writers, there exists a clear range of views on the question of human rights concept. The differing views reflect the academic as well as political debates on the definition and focus of human rights. Somehow some of the views mirror the Cold War scenario and the north-south divide. One can also detect the view that is fit for political expediency agenda of some governments. The varying views, however, can also be seen as results of the different histories of evolution of the idea of human rights. The writers trace the conceptualization of human rights based on the ideas coming from people with dissimilar backgrounds*." HURIGHTS OSAKA – *Human Rights and Cultural Values: A Literature Review*.

[517] Uma das estratégias cogitadas por autores especializados consiste no chamado "*participatory design*", surgido em países nórdicos e no qual se procura promover a participação democrática na construção de parâmetros de design por aqueles que sofreram os possíveis efeitos deletérios dos sistemas em desenvolvimento. NISSENBAUM, Helen – *Values in Technical Design*, p. 3.

[518] Dentre os estudos em torno da matéria, pode ser citado: WONG, Park-Hang – *Cultural Differences as Excuses? Human Rights and Cultural Values in Global Ethics and Governance of AI*.

[519] O texto completo pode ser encontrado em: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence [Acesso em 21/04/21]

[520] Em Portugal, vale referir a Carta Portuguesa de Direitos Humanos na Era Digital, que enaltece o respeito aos direitos fundamentais pela IA, em seu art.º 9.º, cujo teor vale registrar: "Uso da inteligência artificial e de robôs 1 - A utilização da inteligência artificial deve ser orientada pelo respeito dos direitos fundamentais, garantindo um justo equilíbrio entre os princípios da explicabilidade, da segurança, da transparência e da responsabilidade, que atenda às circunstâncias de cada caso concreto e estabeleça processos destinados a evitar quaisquer preconceitos e formas de discriminação. 2 - As decisões com impacto significativo na esfera dos destinatários que sejam tomadas mediante o uso de algoritmos devem ser comunicadas aos interessados, sendo suscetíveis de recurso e

destacam a dignidade humana, a privacidade, a não-discriminação e a liberdade de expressão[521].

No concernente à dignidade humana, na esteira dos ensinamentos de JORGE REIS NOVAIS, não é demais recordar sua "importância essencial não apenas como primeira referência simbólica, de toda ordem constitucional, mas também enquanto princípio de onde decorrem as consequências práticas próprias da irredutível inconstitucionalidade de que enfermam quaisquer violações do princípio"[522].

Constitui o "alfa e ômega", conforme lembra JORGE MIRANDA[523], a "pedra angular" da arquitetura das liberdades constitucionais e dos direitos fundamentais, tal como reconhecidos em boa parte dos Estados de Direito Democráticos. Representa, pois, o "núcleo axiológico das constituições contemporâneas", na expressão de SARLET[524].

Portanto, pode-se ter aí na dignidade da pessoa humana um ponto de convergência para onde devem talvez ser dirigidos os esforços para construção dos necessários consensos quanto aos valores a serem universalmente resguardados na tecnoregulação da IA.

4.3.2.2. Supervisão

A existência de instâncias de supervisão e controle revela-se imprescindível à concretização dos objetivos regulatórios supradelineados.

---

auditáveis, nos termos previstos na lei.3 - São aplicáveis à criação e ao uso de robôs os princípios da beneficência, da não-maleficência, do respeito pela autonomia humana e pela justiça, bem como os princípios e valores consagrados no artigo 2.º do Tratado da União Europeia, designadamente a não discriminação e a tolerância".

[521] "*The use of AI with its specific characteristics (e.g. opacity, complexity, dependency on data, autonomous behaviour) can adversely affect a number of fundamental rights enshrined in the EU Charter of Fundamental Rights ('the Charter'). This proposal seeks to ensure a high level of protection for those fundamental rights and aims to address various sources of risks through a clearly defined risk-based approach. With a set of requirements for trustworthy AI and proportionate obligations on all value chain participants, the proposal will enhance and promote the protection of the rights protected by the Charter: the right to human dignity (Article 1), respect for private life and protection of personal data (Articles 7 and 8), nondiscrimination (Article 21) and equality between women and men (Article 23). It aims to prevent a chilling effect on the rights to freedom of expression (Article 11) and freedom of assembly (Article 12), to ensure protection of the right to an effective remedy and to a fair trial, the rights of defence and the presumption of innocence (Articles 47 and 48), as well as the general principle of good administration. Furthermore, as applicable in certain domains, the proposal will positively affect the rights of a number of special groups, such as the workers' rights to fair and just working conditions (Article 31), a high level of consumer protection (Article 28), the rights of the child (Article 24) and the integration of persons with disabilities (Article 26). The right to a high level of environmental protection and the improvement of the quality of the environment (Article 37) is also relevant, including in relation to the health and safety of people*". Fonte: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence [Acesso em 21/04/21]

[522] NOVAIS, Jorge Reis - A Dignidade da Pessoa Humana: Dignidade e Direitos Fundamentais, p. 20.

[523] MIRANDA, Jorge – Manual de Direito Constitucional, vol. IV, p. 200

[524] *Apud* SANTOS, Coriolano Aurélio de Almeida Camargo & EROUD Aicha – Inteligência artificial e direitos humanos: Uma possível dignidade da pessoa humana digital?.

Tais órgãos deveriam reunir, idealmente, algumas características para assegurar a eficiência na implementação de soluções voltadas para o desenvolvimento e adoção de sistemas desenhados para serem "*human rights friendly*".

O primeiro atributo relevante seria o seu caráter supranacional, em virtude da evidente ubiquidade de boa parte dos sistemas de IA, os quais costumam manter algum nível de conexão com a rede mundial de computadores.

A atuação coordenada de Estados, organização não governamentais, instituições acadêmicas e empresas privadas pode ser decisiva para inclusive para impedir os riscos existenciais resultantes, especialmente, do denominado "*fast takeoff*"[525] ("decolagem rápida")[526].

Entidades internacionais representativas podem assumir algum protagonismo na definição de políticas, instrumentos de "*soft law*" e mesmo criação de observatórios[527] para acompanhar a evolução do tema nos seus países membros.

A esse propósito, em 2021, o Secretário Geral da Organização das Nações Unidas (ONU) apresentou proposta para criar um corpo consultivo multidisciplinar e composto de entidades de diversos setores, cujo objetivo seria o de desenvolver a cooperação internacional quanto à IA [528].

A atuação desse órgão ocorreria em três âmbitos: i) inclusão – construir capacidade global para o desenvolvimento e uso de IA; ii) coordenação – abordar a falta de representação e inclusão em discussões globais; e iii) desenvolvimento de capacidades – empregar a IA para apoiar os objetivos de desenvolvimento sustentável [529].

Em segundo lugar, aos entes de supervisão deveria ser conferida alguma capacidade coercitiva, ainda que por meios indiretos. Aqui seria quase inevitável a mobilização das instituições nacionais.

---

[525] A decolagem rápida consistiria na possibilidade de desenvolvimento ultrarrápido de sistemas de IA dotados de "superinteligência". "*A fast takeoff is one that occurs over the timescale of minutes, hours, or days*". Fonte: https://www.alignmentforum.org/posts/YgNYA6pj2hPSDQiTE/distinguishing-definitions-of-takeoff [Consult. 6 dez. 2022].

[526] "*At around the time risk from superintelligent AI could begin, governments could coordinate to make sure that any risky aspects of AI development are conducted safely. This could involve the creation of a joint international project*". Fonte: https://www.danieldewey.net/fast-takeoff-strategies.html [Consult. 6 dez. 2022].

[527] A Organização para a Cooperação e Desenvolvimento Econômico ou Económico mantém um observatório com esse propósito. Trata-se do OECD AI Policy Observatory (OECD.AI), para avaliar a aderência de sua Recomendação sobre Inteligência Artificial ("*OECD AI Principles*"). Maiores informações podem ser encontradas em: https://oecd.ai/en/about [Consult. 25 nov. 2022].

[528] Fonte: https://www.un.org/techenvoy/content/artificial-intelligence [Consult. 5 dez. 2022]. Para uma visão panorâmica das atividades desenvolvidas pela ONU quanto ao tema da IA: *UN – United Nations Activities on Artificial Intelligence (AI) 2021*.

[529] Fonte: https://www.un.org/techenvoy/content/artificial-intelligence [Consult. 5 dez. 2022].

Por isso mesmo, a proposta regulatória europeia confere aos Estados Membros a incumbência de designar autoridades de supervisão nacional, destinadas a promover a "vigilância" e notificação dos desenvolvedores de sistemas de IA, em casos de infração aos preceitos[530].

Convém salientar que a Espanha anunciou em janeiro de 2022 a instituição de uma entidade dessa natureza, a "*Agencia Estatal de Supervisión de Inteligencia Artificial*" (AESIA)[531].

Dotada de personalidade jurídica própria, a referida agência estará vinculada à "*Secretaría de Estado de Digitalización e Inteligencia Artificial del Ministerio de Asuntos Económicos y Transformación Digital*"[532]. Contaria com capacidade "*inspectora*" e "*sacionadora*", de forma a atuar "*como organismo público con personalidad jurídica pública, patrimonio propio, plena capacidad de obrar y las potestades administrativa, inspectora y sancionadora que se le atribuyan en aplicación de la normativa nacional y europea en relación con el uso seguro y confiable de los sistemas de inteligencia artificial*"[533].

Entre os seus objetivos figuraria o de realizar a supervisão do desenvolvimento e comercialização de sistemas de IA, com especial atenção para aqueles que produzam riscos significativos para a saúde, a segurança e os direitos fundamentais[534].

Em setembro de 2022, deu-se o início do processo para definir a sede da AESIA[535], que já conta com um orçamento inicial de 5 milhões de euros[536].

---

[530] Eis o que prevê o art.º 59º da Proposta de Regulação Europeia: "*Article 59 Designation of national competent authorities 1. National competent authorities shall be established or designated by each Member State for the purpose of ensuring the application and implementation of this Regulation. National competent authorities shall be organised so as to safeguard the objectivity and impartiality of their activities and tasks. 2. Each Member State shall designate a national supervisory authority among the national competent authorities. The national supervisory authority shall act as notifying authority and market surveillance authority unless a Member State has organisational and administrative reasons to designate more than one authority*".

[531] Fonte: https://www.lexology.com/library/detail.aspx?g=d2d01036-bbdf-451a-b3e7-485f51148b76 [Consult. 2 dez. 2022]. Note-se que diversas organizações da sociedade civil espanhola subscreveram uma carta aberta com o objetivo de exigir que a nova agência priorize os direitos humanos e a justiça social. O manifesto pode ser encontrado em: https://www.statewatch.org/news/2022/september/spain-new-artificial-intelligence-agency-must-put-human-rights-and-social-justice-at-the-centre/ [Consult. 2 dez. 2022].

[532] Fonte: https://elderecho.com/creada-agencia-supervision-inteligencia-artificial [Consult. 2 dez. 2022].

[533] Fonte: https://elderecho.com/creada-agencia-supervision-inteligencia-artificial [Consult. 2 dez. 2022].

[534] Fonte: https://elderecho.com/creada-agencia-supervision-inteligencia-artificial [Consult. 2 dez. 2022].

[535] Fonte: https://portal.mineco.gob.es/es-es/comunicacion/Paginas/agencia-espa%C3%B1ola-de-supervisi%C3%B3n-de-inteligencia-artificial.aspx [Consult. 2 dez. 2022]. O Real Decreto 209/2022 em que a matéria é disciplinada encontra-se disponível em: https://www.boe.es/buscar/act.php?id=BOE-A-2022-4630 [Consult. 2 dez. 2022].

[536] Fonte: https://protecciondata.es/agencia-espanola-supervision-inteligencia-artificial/ [Consult. 2 dez. 2022]. A Ley 22/2021 que define os aspectos orçamentários pode ser encontrada em: https://www.boe.es/buscar/act.php?id=BOE-A-2022-4630 [Consult. 2 dez. 2022].

Na mesma linha, em novembro de 2022, a Holanda oficializou seu intuito de implementar provisões da proposta regulatória europeia, mesmo antes de entrar em vigor, formalmente. Nesse contexto, a partir de janeiro de 2023, deve instalar um corpo de supervisão destinado a impor um conjunto de novas regras concernentes a aspectos como as obrigações de transparência dos sistemas[537].

O terceiro elemento seria a natureza idealmente preventiva da atuação das instâncias supervisoras, mediante o uso de técnicas de "avaliação de impacto algorítmico"[538]. Seria imperioso, nessa perspectiva, exigir dos desenvolvedores que realizassem (e compartilhassem) essa análise apriorística, i.e., procedessem a uma rigorosa varredura dos riscos de vulneração de direitos fundamentais, antes da implementação dos sistemas.

O quarto e último traço relevante da supervisão dos sistemas de IA seria o forte emprego de instrumentos tecnológicos, particularmente aqueles artificialmente inteligentes[539]. O que se propõe aqui é a utilização da própria IA para acompanhar, monitorar e mesmo intervir para neutralizar ou minimizar os riscos já abordados[540].

---

[537] Fonte: https://www.euractiv.com/section/digital/news/once-bitten-netherlands-wants-to-move-early-on-algorithm-supervision/ [Consult. 2 dez. 2022].

[538] A esse respeito, ressalta SELBST: "*Scholars and advocates have proposed algorithmic impact assessments ("AIAs") as a regulatory strategy for addressing and correcting algorithmic harms. An AIA-based regulatory framework would require the creator of an algorithmic system to assess its potential socially harmful impacts before implementation and create documentation that can be used later for accountability and future policy development. In practice, an impact assessment framework relies on the expertise and information to which only the creators of the project have access. It is therefore inevitable that technology firms will have an amount of practical discretion in the assessment, and willing cooperation from firms is necessary to make the regulation work. But a regime that relies on good-faith partnership from the private sector also has strong potential to be undermined by the incentives and institutional logics of the private sector. This Article argues that for AIA regulation to be effective, it must anticipate the ways that such regulation will be filtered through the private sector institutional environment*.". SELBST, Andrew D. – *An Institutional View of Algorithmic Impact Assessments*.

[539] Nessa linha já existem soluções como a plataforma privada SECLEA desenvolvida com objetivos semelhantes aos descritos, conforme se extrai da descrição contida em seu sítio na Internet: "*AI risk management is a continuous process that is critical to machine learning and deep learning applications at all stages of their lifecycle for both ensuring the applications behave in a responsible manner and regulatory compliance. Seclea Platform provides real-time monitoring of risk mitigations actions related to machine learning and deep learning algorithms at every lifecycle stage. Tracking both the human (data scientists, machine learning engineers, AI project leads, auditors, etc.) and machine (data and algorithm) activities gives your company the most comprehensive picture of your AI application's risk posture*". Fonte: https://www.seclea.com/seclea_platform/ [Consult. 19 dez. 2022].

[540] Uma perspectiva interessante concerne ao uso tecnoregulatório da IA para impedir que agentes artificialmente inteligentes provoquem danos a direitos fundamentais. Um exemplo nessa direção é a identificação, via ferramentas de IA, de pessoas vulneráveis (crianças, idosos, etc.) a partir do cruzamento de dados como o padrão de toques na tela de smartphones para garantir uma proteção adicional a tais indivíduos diante de ameaças *online*, em grande parte movidas por IA, atualmente. A esse respeito, vide: ZACCAGNINO, Rocco & CAPO, Carmine & GUARINO, Alfonso & LETTIERI, Nicola & MALANDRINO, Delfina – *Techno-regulation and intelligent safeguards Analysis of touch gestures for online child protection*.

CONCLUSÕES

1. Conclusões artificiais

"Em geral, quando se trata de regulamentação de IA e direitos humanos, há alguns pontos importantes a se considerar. Primeiro, é importante reconhecer que sistemas de IA podem ter um impacto profundo na vida das pessoas, tanto positivamente quanto negativamente. Isso significa que é crucial que governos e outras organizações considerem cuidadosamente as possíveis consequências do uso de IA e coloquem em prática medidas de proteção adequadas para proteger os direitos das pessoas.

Um dos principais desafios na regulamentação de IA é o fato de ser uma tecnologia em constante evolução. Isso significa que as regulamentações devem ser flexíveis o suficiente para se adaptar a novos desenvolvimentos, ao mesmo tempo em que garantem a proteção dos direitos das pessoas. É também importante que os reguladores consultem uma ampla gama de stakeholders, incluindo especialistas em IA e direitos humanos, para garantir que as regulamentações sejam eficazes e adequadas.

Outro aspecto importante da regulamentação de IA é a necessidade de garantir transparência e responsabilidade. Isso significa que organizações que utilizam IA devem ser obrigadas a fornecer explicações claras e concisas sobre o funcionamento de seus sistemas e devem ser responsabilizadas por quaisquer impactos negativos que possam ter nos direitos das pessoas.

Em geral, a chave para uma regulamentação de IA bem-sucedida é equilibrar a possibilidade de inovação e progresso com a proteção dos direitos das pessoas. Isso exigirá muito pensamento e consideração, bem como um diálogo contínuo entre reguladores, indústria e outros stakeholders. Portanto, para concluir, é essencial que governos e outras organizações considerem cuidadosamente as possíveis consequências do uso de IA e coloquem em prática medidas de proteção adequadas para proteger os direitos das pessoas".[541]

[541] No original: "[…]. *In general, when it comes to AI regulation and human rights, there are a few key points to consider. First, it's important to recognize that AI systems can have a profound impact on people's lives, both positively and negatively. This means that it's crucial for governments and other organizations to carefully consider the potential consequences of using AI, and to put in place appropriate safeguards to protect people's rights. One of the key challenges in regulating AI is the fact that it is a rapidly evolving technology. This means that regulations must be flexible enough to adapt to new developments, while still ensuring that people's rights are protected. It's also important for regulators to consult with a wide range of stakeholders, including experts in AI*

Não tivessem sido incluídas as aspas no texto acima (propositalmente sem o recuo exigido pelas normas técnicas), talvez se pudesse concebê-lo como o fruto do esforço criativo do autor da presente dissertação.

Embora empregue uma linguagem um tanto informal e mantenha-se em um certo nível de superficialidade, trata-se de um feito impressionante se considerada a circunstância de que resultou de uma elaboração integral de um agente artificialmente inteligente.

As conclusões supratranscritas de resposta (propositalmente sem aspas e sem a devida formatação própria às citações) à pergunta formulada por este autor[542] a um sistema de geração de textos em linguagem natural[543] desenvolvido pela empresa "OpenAI"[544].

Fruto de um mero exercício de interação mais "direta" com a IA, o objetivo da exposição supra foi o de apresentar o resultado de uma sintética compilação de aspectos relevantes a respeito do tema objeto da investigação conduzida por este investigador, a partir de uma avaliação feita pelo "destinatário" primário da regulação aqui proposta.

2. Conclusões pessoais

No panteão dos clássicos mitos gregos figura em destaque o de Prometeu (do grego Προμηθεύς, "antevisão"), esculpido de forma indelével por autores como Hesíodo, Ésquilo, Ovídio, entre outros. A trama, como se sabe, desenrola-se a partir da subtração do fogo sagrado de Héstia pelo titã, que o entregou aos homens, crime pelo qual foi condenado a

---

and human rights, in order to ensure that regulations are effective and appropriate. Another key aspect of AI regulation is the need to ensure transparency and accountability. This means that organizations using AI should be required to provide clear and concise explanations of how their systems work, and should be held accountable for any negative impacts they may have on people's rights. Overall, the key to successful AI regulation is to strike a balance between allowing for innovation and progress, while at the same time protecting people's rights. This will require careful thought and consideration, as well as ongoing dialogue between regulators, industry, and other stakeholders. So, to conclude, it is essential for governments and other organizations to carefully consider the potential consequences of using AI, and to put in place appropriate safeguards to protect people's rights". Fonte: https://chat.openai.com/chat [Consult. 08 dez. 2022]

[542] Eis o teor do pedido: "*Please, I would like your help to make a conclusion for my paper on "Human righsts by design" about AI regulation*". Fonte: https://chat.openai.com/chat [Consult. 08 dez. 2022]

[543] Eis o texto autodescritivo gerado pelo sistema "*I am a large language model trained by OpenAI. I am able to generate text based on the input that I receive from users like you. I am not a person, but rather a computer program designed to process and generate natural language text. I do not have my own thoughts or opinions, and I can only generate text based on the information that I have been trained on. I am here to assist you with your questions and provide you with information to the best of my ability*". Fonte: https://chat.openai.com/chat [Consult. 08 dez. 2022]

[544] "OpenAI's mission is to ensure that artificial general intelligence (AGI)—by which we mean highly autonomous systems that outperform humans at most economically valuable work—benefits all of humanity. We will attempt to directly build safe and beneficial AGI, but will also consider our mission fulfilled if our work aids others to achieve this outcome". Fonte: https://openai.com/about/ [Consult. 08 dez. 2022]

permanecer atado a uma rocha no monte Cáucaso, sendo fustigado por uma águia a devorar seu regenerável fígado todos os dias, até ser libertado por Hércules após ter concluído seus doze trabalhos.

Com certa frequência a imagem é evocada ao se descrever o salto que a humanidade poderia vir a dar ao desenvolver artefatos ou, talvez se possa afirmar, "seres" tecnológicos para os quais transferiria uma das mais importantes potências da alma, a inteligência.

Trata-se, de fato, de um momento disruptivo que poderia romper o equilíbrio autopoiético do movimento pendular característico da evolução histórica humana.

A inteligência artificial figura entre as tecnologias com maior potencial de ruptura do curso da história do homem, ao inaugurar novas perspectivas de transformação do mundo natural.

Seu alcance transcenderia, virtualmente, as fronteiras descritas até aqui pela inteligência "natural", alcançando todas as dimensões da atividade humana, além de descortinar horizontes inéditos.

Embora não conte com uma definição técnica precisa e científica, a inteligência artificial pode ser vista como um conjunto de técnicas, conhecimentos, princípios e conceitos que permitem a transferência de parte das capacidades cognitivas humanas a agentes “sintéticos”, de forma a habilitá-los para a resolução de problemas matematicamente definíveis.

Sua evolução foi o resultado de uma trajetória não linear com primaveras, verões, outonos e invernos, não necessariamente nessa ordem. Os potenciais da tecnologia têm sido objeto ora de elevado otimismo, ora de pessimismo em demasia.

Há, contudo, alguma convergência em torno dos riscos presentes e futuros relativos aos seus impactos disruptivos quanto a diversos aspectos da vida humana, em particular no tocante aos direitos fundamentais cuja vulnerabilidade pode ser reconhecida em diversas dimensões. Vida, integridade física, dignidade, liberdade e igualdade são algumas das garantias universais direta e indiretamente suscetíveis a danos resultantes da implementação de soluções de IA.

O ritmo vertiginoso e a imprevisibilidade dos rumos a serem tomados nesse campo tecnológico representam alguns dos desafios para conter tais ameaças. Além disso, algumas das características intrínsecas à maior parte dos sistemas de IA também dificultam a criação de barreiras de contenção eficientes.

A sua opacidade, estrutura correlacional, complexidade e autonomia têm sido alguns dos aspectos presentes no panorama hodierno a desafiar a inteligência legiferante.

Entretanto, mesmo no horizonte de curto prazo já podem ser percebidos os primeiros frutos do esforço regulatório dedicado ao tema, em uma profusão de instrumentos sobretudo de "*soft law*", mas também de "*hard law*", esses últimos ainda de forma fragmentada, em regra. Como exceção situa-se, nesse último conjunto, a proposta europeia de regulação abrangente da matéria.

A centralidade do homem emerge como um parâmetro norteador presente na maior parte das tentativas conduzidas por entes públicos e privados para conter os potenciais deletérios da inteligência das máquinas, com uma especial atenção para o superprincípio da dignidade da pessoa humana.

Nessa perspectiva, como instrumento para alinhar valores humanos e metas perseguidas por agentes artificialmente inteligentes, a abordagem "*human rights by design*" parece uma via tecnoregulatória de elevado potencial.

Com efeito, revela-se promissora a criação de "barreiras de contenção" normativas por meio da introdução de "travas" desenhadas na própria arquitetura dos sistemas, de modo a inviabilizar o cometimento de agressões aos direitos humanos de modo quase absoluto.

Sua eficiência depende em grande medida da superação de ingentes desafios epistemológicos e "implementacionais", entre os quais se avulta a construção de consenso em torno do alcance semântico dos direitos humanos, além da existência de mecanismos de supervisão dotados de supranacionalidade e mínima capacidade coercitiva.

O êxito das empreitadas nessa direção pode levar à salvaguarda da dignidade humana em um cenário de ruptura significativa do equilíbrio autopoiético entre vetores de forças a apontarem em direções diversas e com desdobramentos imprevistos.

O fogo de Prometeu trouxe a luz da inteligência e o calor das ciências aos homens a descortinarem o céu como horizonte de sua transcendência, mas também permitiu lançar muitos às frias sombras da ignorância ou à desoladora geena do materialismo.

O que esperar dos novos portadores da chama cognitiva, máquinas forjadas à imagem e semelhança de seus criadores? A resposta a indagação dependerá do êxito em se fazer incorporar nesses novos agentes os traços axiológicos que distinguem a espécie humana e lhe conferem sentido existencial.

FONTES

1. Bibliográficas

AGGARWAL, Alok - ***The Birth of AI and The First AI Hype Cycle*** [Em linha]. [Consult. 24 dez. 2018] Disponível em https://www.kdnuggets.com/2018/02/birth-ai-first-hype-cycle.html

AIZENBERG Evgeni & HOVEN Jeroen van den - ***Designing for human rights in AI***. [Em linha]. [Consult. 04 out. 2022] Disponível em: *https://journals.sagepub.com/doi/epub/10.1177/2053951720949566*

ALEXANDRINO, José Melo – **Dez breves apontamentos sobre a Carta Portuguesa de Direitos Humanos na Era Digital** [Em linha]. [Consult. 15 dez. 2022] Disponível em https://www.icjp.pt/sites/default/files/papers/dez_breves_apontamentos_sobre_a_carta_portuguesa.pdf

ALFONSECA, Manuel & CEBRIAN, Manuel & ANTA, Antonio Fernández & COVIELLO, Lorenzo & ABELIUK, Andrés & RAHWAN, Iyada – ***Superintelligence Cannot be Contained: Lessons from Computability Theory*** [Em linha]. [Consult. 21 abr. 2022] Disponível em: https://arxiv.org/abs/1607.00913

AMODEI, Dario & OLAH, Chris & SCHULMAN, John & STEINHARDT, Jacob & CHRISTIANO, Paul & MANÉ, Dan – ***Concrete Problems in AI Safety*** [Em linha]. [Consult. 15 jul. 2019] Disponível em: https://arxiv.org/abs/1606.06565

ANDRADE, Francisco & NOVAIS, Paulo & NEVES, José - ***Issues on Intelligent Electronic Agents and Legal Relations*** [Em linha]. [Consult. 11 nov. 2019] Disponível em: https://www.researchgate.net/publication/228598601_Issues_on_intelligent_electronic_agents_and_legal_relations

ANYOHA, Rockwell - ***Can Machines Think?*** [Em linha]. [Consult. 8 jan. 2020] Disponível em: http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/

AZIMOV, Isaac - ***I, Robot***. Gnome Press, 1950

AZOULAY, Audrey - ***Towards an Ethics of Artificial Intelligence*** [Em linha]. [Consult. 17 jan. 2020] Disponível em: https://www.un.org/en/chronicle/article/towards-ethics-artificial-intelligence

BALAGURUNATHAN, Yoganand & MITCHELL, Ross & EL NAQA, Issam - ***Requirements and reliability of AI in the medical context*** [Em linha]. [Consult. 17 jan.

2020] Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S1120179721001137

BARLOW, John Perry – ***A Declaration of the Independence of Cyberspace*** [Em linha]. [Consult. 19 jul. 2022] Disponível em: https://www.eff.org/cyberspace-independence

BAROCAS, Solon and SELBST, Andrew D. - ***Big Data's Disparate Impact*** (2016). 104 California Law Review 671 (2016). [Em linha]. [Consult. 28 dez. 2018] Disponível em Available at http://dx.doi.org/10.2139/ssrn.2477899

BAUMAN, Zygmunt – **Modernidade líquida**. Rio de Janeiro: Editora Zahar, 2001

BAYAMLIOĞLU, Emre & LEENES, Ronald - ***The 'rule of law' implications of data-driven decision- making: a techno-regulatory perspective***. LAW, INNOVATION AND TECHNOLOGY, 2018, VOL. 10, NO. 2, 295–313.

BELLIA, Anthony J. – ***Contracting with Electronic Agents*** [Em linha]. [Consult. 03 dez. 2020] Disponível em: https://scholarship.law.nd.edu/law_faculty_scholarship/101/

BELLOVIN, Steven M. & HUTCHINS, Renée M. & JEBARA, Tony & ZIMMECK, Sebastian – ***When Enough is Enough: Location Tracking, Mosaic Theory, and Machine Learning***. [Em linha]. [Consult. 05 nov. 2021] Disponível em: https://digitalcommons.law.umaryland.edu/fac_pubs/1375/

BOSTROM, Nick - ***Ethical Issues in Advanced Artificial Intelligence*** [Em linha]. [Consult. 11 jan. 2018] Disponível em https://nickbostrom.com/ethics/ai.pdf

– ***Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards***. [Em linha]. [Consult. 24 jun. 2021] Disponível em: https://www.nickbostrom.com/existential/risks.html.

– ***Superintelligence – Paths, Dangers, Strategies***. Londres: Oxford University Press, 2014.

BROWNSWORD, Roger - ***What the World Needs Now: Techno-Regulation,Human Rights and Human Dignity*** [Em linha]. [Consult. 24 dez. 2018] Disponível em: https://kclpure.kcl.ac.uk/portal/en/publications/what-the-world-needs-now-technoregulation-human-rights-and-human-dignity(3cc2cf53-bf1f-4a53-b6e9-ec7d72bb3a2c)/export.html

BUKHARI, S. M. A. H. & KHAN, M. U. S. & KHAN, T. & ALI, M. - ***Social media, AI, and political affiliation*** [Em linha]. [Consult. 22 mar. 2022] Disponível em: https://ieeexplore.ieee.org/document/9709624

CALICE, Mikhaila N. & BAO, Luye & SCHEUFELE, Dietram A. – ***Polarized platforms? How partisanship shapes perceptions of "algorithmic news bias"*** [Em linha]. [Consult.

23 mar. 2022] Disponível em: https://journals.sagepub.com/doi/full/10.1177/14614448211034159

CANOTILHO, José Joaquim Gomes – **Sobre a indispensabilidade de uma Carta de Direitos Fundamentais Digitais da União Europeia** [Em linha]. [Consult. 12 abr. 2022] Disponível em: **https://revista.trf1.jus.br/trf1/article/view/17**

CASTELLS, Manuel - **A Sociedade em Rede. A Era da Informação: Economia, Sociedade e Cultura - Volume 1.** Lisboa: Fundação Calouste Gulbenkian, 2003.

CHRISTIAN, Brian – ***The Alignment Problem – How Can Artificial Intelligence Learn Human Values***. Inglaterra, Londres: Atlantic Books, 2021.

CITRON, Danielle Keats & PASQUALE, Frank – ***The Scored Society: Due Process for Automated Predictions*** [Em linha]. [Consult. 2 fev. 2022] Disponível em: https://digitalcommons.law.uw.edu/wlr/vol89/iss1/2/

CLIFFORD LAW - ***The Dangers of Driverless Cars*** – *In The National Law Review, January, 27, 2022, Volume XII, Number 27*. [Em linha]. [Consult. 27 jan. 2022] Disponível em: https://www.natlawreview.com/article/dangers-driverless-cars

CORTEZ, Nathan - ***Regulating Disruptive Innovation*** [Em linha]. [Consult. 13 nov. 2022] Disponível em: https://www.jstor.org/stable/24119940

COWGILL, Bo - ***Bias and Productivity in Humans and Algorithms: Theory and Evidence from Resume´s Screenin*** [Em linha]. [Consult. 8 jun. 2022] Disponível em: *https://conference.iza.org › cowgill_b8981*

DAMIANO, Jean-Pierre – ***Biomimétisme, intelligence artificielle et robotique. Applications de l'intelligence en essaim et questions d'éthique et de droit***. [Em linha]. [Consult. 4 out. 2022] Disponível em: https://hal.science/hal-03701371

DANKS, David & LONDON, Alex John – ***Algorithmic Bias in Autonomous Systems***. [Em linha]. [Consult. 5 out. 2022] Disponível em: https://www.researchgate.net/publication/318830422_Algorithmic_Bias_in_Autonomous_Systems

DE STEFANO, Valerio e TAES, Simon - ***Algorithmic management fits into a broader context of the implementation of technological tools and digitalised supervision systems aimed at governing the workforce***. [Em linha]. [Consult. 26 jun. 2021] Disponível em: https://www.etui.org/publications/algorithmic-management-and-collective-bargaining

EGGERS, William D. & TURLEY, Mike & KISHNANI, Pankaj – ***The future of regulation Principles for regulating emerging technologies*** [Em linha]. [Consult. 6 dez. 2022] Disponível em: https://www2.deloitte.com › dam › insights › articles

ELIOT, Lance B. – ***Using AI As A Techno-Regulatory Enforcement Tool***. [Em linha]. [Consult. 25 nov. 2022] Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3998246

ERNST, Ekkehardt & MEROLA, Rossana & SAMAAN, Daniel – ***Economics of artificial intelligence: Implications for the future of work*** [Em linha]. [Consult. 29 jun. 2021] Disponível em: https://doi.org/10.2478/izajolp-2019-0004.

FAST, Ethan & HORVITZ, Eric - ***Long-Term Trends in the Public Perception of Artificial Intelligence*** [Em linha]. [Consult. 15 mar. 2021] Disponível em: https://arxiv.org/pdf/1609.04904.pdf

FITCH, Andy – ***Letter from Utopia: Talking to Nick Bostrom***. [Em linha]. [Consult. 10 out. 2022] Disponível em: https://blog.lareviewofbooks.org/interviews/letter-utopia-talking-nick-bostrom/

FJELD, Jessica & ACHTEN, Nele & HILLIGOSS, Hannah & NAGY, Adam & SRIKUMAR, Madhulika – ***Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI***. [Em linha]. [Consult. 20 dez. 2022] Disponível em: https://dash.harvard.edu/handle/1/42160420

FLANAGAN, Mary & HOWE, Daniel C. & NISSENBAUM, Helen – ***Embodying Values in Technology – Theory and Practice*** [Em linha]. [Consult. 7 nov. 2022] Disponível em: https://www.researchgate.net/publication/288448003_Embodying_values_in_technology_Theory_and_practice

FLOWERS, Johnathan Charles - ***Strong and Weak AI: Deweyan Considerations*** [Em linha]. [Consult. 17 fev. 2021] Disponível em: http://ceur-ws.org/Vol-2287/paper34.pdf

FRIEDMAN, Batya – ***Human Values and the Design of Computer Technology***. Stanford: Center for the Study of Language and Information, 1997.

GALL, Richard – *Machine Learning Explainability vs Interpretability: Two concepts that could help restore trust in AI* [Em linha]. [Consult. 6 jan. 2023] Disponível em: *https://www.kdnuggets.com/2018/12/machine-learning-explainability-interpretability-ai.html*

GONÇALVES, Maria Eduarda & GAMEIRO, Maria – ***Hard Law, Soft Law and Self-regulation: Seeking Better Governance for Science and Technology in the EU*** [Em linha]. [Consult. 18 jul. 2022] Disponível em: https://www.researchgate.net/publication/272351073_Hard_Law_Soft_Law_and_Self-regulation_Seeking_Better_Governance_for_Science_and_Technology_in_the_EU/link/54e227fd0cf29666379579dd/download

GRIEMAN, Keri - *Hard Drive Crash* ***An Examination of Liability for Self-Driving Vehicles*** [Em linha]. [Consult. 12 fev. 2021] Disponível em: *https://www.jipitec.eu/issues/jipitec-9-3-2018/4806*

GWERN - ***Complexity no Bar to AI***. Em linha]. [Consult. 21 mar. 2021] Disponível em: https://www.gwern.net/Complexity-vs-AI

HAGEMANN, Ryan & SKEES, Jennifer Huddleston & THIERER, Adam – ***Soft Law For Hard Problems: The Governance Of Emerging Technologies In An Uncertain Future Intelligence*** [Em linha]. [Consult. 11 jul. 2019] Disponível em: https://ssrn.com/abstract=3118539

HAO, Karen – ***This is how AI bias really happens—and why it's so hard to fix*** [Em linha]. [Consult. 26 ago. 2020] Disponível em: https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/

HARTSON, Rex & PYLA, Pardha – ***History and Origins of Participatory Design***. [Em linha]. [Consult. 14 out. 2022] Disponível em: https://www.sciencedirect.com/topics/computer-science/participatory-design

HENDRY, David G. & FRIEDMAN, Batya & BALLARD, Stephanie – ***Value sensitive design as a formative framework***. [Em linha]. [Consult. 04 out. 2022] Disponível em: https%3A%2F%2Fd-nb.info%2F1231350873%2F34&usg=AOvVaw20GwDXbTmiqy2LzMuuUyw5

HILDEBRANDT, Mireille - ***Smart Technologies and the End(s) of Law***. Cheltenham (UK): Edward Elgar Publishing. ISBN 9781849808774

HORNE, Benjamin D. & NEVO, Dorit & ADALI, Sibel & MANIKONDA, Lydia & ARRINGTON, Clare – ***Tailoring heuristics and timing AI interventions for supporting news veracity assessments***. [Em linha]. [Consult. 05 nov. 2021] Disponível em: https://www.sciencedirect.com/science/article/pii/S2451958820300439

HSU, Jeremy – ***AI Recruiting Tools Aim to Reduce Bias in the Hiring Process*** [Em linha]. [Consult. 23 abr. 2020] Disponível em: https://spectrum.ieee.org/ai-tools-bias-hiring

HUME, David – ***A Treatise of Human Nature***. Londres: Oxford University Press, 1978.

JHA, Susmit & RAJ, Sunny & FERNANDES, Steven - ***Attribution-Based Confidence Metric For Deep Neural Networks*** [Em linha]. [Consult. 12 ago. 2020] Disponível em:

JOHNSON, Deborah G. & VERDICCHIO, Mario - ***AI, agency and responsibility: the VW fraud case and beyond*** [Em linha]. [Consult. 7 set. 2021] Disponível em:

JORDAN, Michael – ***Artificial Intelligence—The Revolution Hasn't Happened Yet*** [Em linha]. [Consult. 25 mai. 2020] Disponível em: https://papers.nips.cc/paper/2019/hash/bc1ad6e8f86c42a371aff945535baebb-Abstract.html

KAPLAN, Andreas & HAENLEIN, Michael - ***Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence***. Business Horizons, Volume 62, Issue 1, January–February 2019

KERR, Ian - ***The Arrival of Artificial Intelligence and "The Death of Contract"*** [Em linha]. [Consult. 27 out. 2020] Disponível em: https://techlaw.uottawa.ca/news/arrival-artificial-intelligence-and-death-contract

KISSINGER, Henry A.; SCHMIDT, Eric & HUTTENLOCHER, Daniel – **A Era da Inteligência Artificial**. Portugal, Alfragide: Dom Quixote, 2021.

KLEINSTEUBER, Hans J. - ***Self-regulation, Co-regulation, State Regulation – The Internet between Regulation and Governance***. [Em linha]. [Consult. 10 set. 2022] Disponível em: www.osce.org%2Ffiles%2Ff%2Fdocuments%2F2%2Fa%2F13844.pdf&usg=AOvVaw0Pq_8xir-Li1tDvie8sqxj

KOTIN, Timothy - ***Artificial intelligence and data analytics can unlock new economic opportunities*** [Em linha]. [Consult. 21 abr. 2022] Disponível em: https://www.oecd-forum.org/posts/in-my-view-artificial-intelligence-and-data-analytics-can-unlock-new-economic-opportunities?badge_id=649-digital-inclusion

KRIEBITZ, Alexander & LÜTGE, Christoph - ***Artificial Intelligence and Human Rights: A Business Ethical Assessment*** [Em linha]. [Consult. 10 ago. 2022] Disponível em: https://www.cambridge.org/core/journals/business-and-human-rights-journal/article/abs/artificial-intelligence-and-human-rights-a-business-ethical-assessment/33D07AB42FC76A4BA49B03F600186E1B

KROETZ, Flávia Saldanha - ***Between global consensus and local deviation: a critical approach on the universality of human rights, regional human rights systems and cultural diversity*** [Em linha]. [Consult. 21 nov. 2022] Disponível em: https://www.scielo.br/j/rinc/a/6DKr5ZGykR7NyZ6Hb3S7ptv/?lang=en

LADER, Bhagirath Kumar - ***Some Thoughts on Artificial Intelligence Singularity*** [Em linha]. [Consult. 15 mar. 2021] Disponível em: https://medium.datadriveninvestor.com/some-thoughts-on-artificial-intelligence-singularity-3f16db2ae8ae

LEENDERS, Gijs – ***The Regulation Of Artificial Intelligence – A Case Study of the Partnership on AI***. [Em linha]. [Consult. 23 set. 2022] Disponível em: https://becominghuman.ai/the-regulation-of-artificial-intelligence-a-case-study-of-the-partnership-on-ai-c1c22526c19f

LEENES, Ronald E - ***Framing Techno-Regulation: An Exploration of State and Non-State Regulation by Technology*** [Em linha]. [Consult. 23 dez. 2018] Disponível em https://www.researchgate.net/publication/233661160_Framing_Techno-Regulation_An_Exploration_of_State_and_Non-State_Regulation_by_Technology

LENARDON, João Paulo de Almeida - ***The Regulation of Artificial Intelligence*** [Em linha]. [Consult. 20 jul. 2019] Disponível em https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=2ahUKEwjk3oz4hpDnAhWOILkGHdGuDHoQFjABegQICRAB&url=http%3A%2F%2Farno.uvt.nl%2Fshow.cgi%3Ffid%3D142832&usg=AOvVaw1obeDqa0KiO_3IYni3qLGk

LESSIG, Lawrence - ***Code and other Laws of Cyberspace***. New York: Basic Books, 1999.

- ***Code Version 2.0***. New York: Basic Books, 2006.

LEVY, Pierre - ***Intelligence artificielle et sciences humaines*** [Em linha]. [Consult. 19 dez. 2019] Disponível em: https://pierrelevyblog.com/2014/10/08/intelligence-artificielle-et-sciences-humaines/

LEWIS, Tanya & WRITER, Staff - ***Brain Says Guilty! Neural Imaging May Nab Criminals***. [Em linha]. [Consult. 5 jan. 2019] Disponível em: https://www.livescience.com/37091-brain-imaging-in-the-courtroom.html

LODDER, Arno R. & ZELEZNIKOW, John - ***Artificial Intelligence and Online Dispute Resolution.*** [Em linha]. [Consult. 29 dez. 2019] Disponível em: https://www.mediate.com/articles/ODRTheoryandPractice4.cfm

LOREGGIA, Andrea & MATTEI, Nicholas & QUINTARELLI, Stefano – ***Artificial Intelligence Research for Fighting Political Polarisation: A Research Agenda*** [Em linha]. [Consult. 9 nov. 2022] Disponível em: https://ceur-ws.org › Vol-2781 › paper2

MANTELLO, Peter & HO, Manh-Tung & NGUYEN, Minh-Hoang & VUONG, Quan-Hoang – ***Bosses without a heart: socio-demographic and cross-cultural determinants of attitude toward Emotional AI in the workplace***. [Em linha]. [Consult. 24 jun. 2022] Disponível em: https://link.springer.com/article/10.1007/s00146-021-01290-1

MARANHÃO, Juliano – ***A pesquisa em inteligência artificial e Direito no Brasil***. [Em linha]. [Consult. 15 jul. 2022] Disponível em: https://www.conjur.com.br/2017-dez-09/juliano-maranhao-pesquisa-inteligencia-artificial-direito-pais

MARCHANT, E. Gary – ***Addressing the Pacing Problem*** [Em linha]. [Consult. 12 ago. 2022] Disponível em: https://www.researchgate.net/publication/303158895_Addressing_the_Pacing_Problem

MARSDEN Christopher T. – ***Co- and Self-regulation in European Media and Internet Sectors: The Results of Oxford University's Study*** [Em linha]. [Consult. 10 set. 2022] Disponível em: www.osce.org%2Ffiles%2Ff%2Fdocuments%2F2%2Fa%2F13844.pdf&usg=AOvVaw0Pq_8xir-Li1tDvie8sqxj

MARWALA, Tshilidzi - ***Causality, Correlation and Artificial Intelligence for Rational Decision Making***. New Jersey: World Scientific, 2015.

MCCARTHY, John - ***What is Artificial Intelligence?*** [Em linha]. [Consult. 8 out. 2020] Disponível em: http://jmc.stanford.edu/articles/whatisai.html

MCCARTHY, John; MINSKY, Marvin; ROCHESTER, Nathan; SHANNON, Claude - ***A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*** (31 August 1955) [Em linha]. [Consult. 23 dez. 2018] Disponível em http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

MIRANDA, Jorge – **Manual de Direito Constitucional**, vol. IV, 4ª ed., Coimbra: Coimbra Editora, 2008.

MODGIL, Sachin & SINGH, Rohit Kumar & GUPTA, Shivam & DENNEHY, Denis – ***A Confirmation Bias View on Social Media Induced Polarisation During Covid-19*** [Em linha]. [Consult. 19 set. 2022] Disponível em: https://link.springer.com/article/10.1007/s10796-021-10222-9

MOSES, Lyria Bennett – ***How to Think about Law, Regulation and Technology: Problems with 'Technology' as a Regulatory Target*** [Em linha]. [Consult. 03 out. 2022] Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2464750

MURRAY, Andrew & SCOTT, Colin - ***Controlling the New Media: Hybrid Responses to New Forms of Power*** *(2002) Modern Law Review 491* [Em linha]. [Consult. 23 dez. 2018] Disponível em *https://www.jstor.org/stable/1097592*

NISSENBAUM, Helen – ***Values in Technical Design***. [Em linha]. [Consult. 05 out. 2022] Disponível em: https://nissenbaum.tech.cornell.edu

NOVAIS, Jorge Reis – **A Dignidade da Pessoa Humana: Dignidade e Direitos Fundamentais**. Coimbra: Almedina, 2015.

NOVAIS, Paulo & FREITAS, Pedro Miguel - **Inteligência Artificial e Regulação de algoritmos** [Em linha]. [Consult. 03 jul. 2019] Disponível em

https%3A%2F%2Fwww.sectordialogues.org%2Fdocumentos%2Fproyectos%2Fadjuntos%2F49f7d3_Intelig%25C3%25AAncia%2520Artificial%2520e%2520Regula%25C3%25A7%25C3%25A3o%2520de%2520Algoritmos.pdf&usg=AOvVaw1gxE3VYFvi6vI57y5e9pVz

O´NEIL, Cathy – ***Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.*** Nova Iorque: Crown, 2016

PAGALLO, Ugo & BAYAMLIOĞLU, Emre & CUIJPERS, Colette& DURANTE, Massimo & LIMA, Marcelo Padua & MAGRANI, Eduardo & ROSINA, Mônica Steffen Guise & TRIPATHY, Sunita & WEILL, Rivka - ***New Technologies and Law: Global Insights on the Legal Impacts of Technology, Law as Meta-Technology and Techno Regulation*** [Em linha]. [Consult. 26 dez. 2018] Disponível em http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html. http://www.lawandtechnology.net/project/

PAULA, Gáudio Ribeiro de – **A Proposta Europeia para regulação da inteligência artificial** [Em linha]. [Consult. 23 abr. 2021] Disponível em: https://www.migalhas.com.br/depeso/344224/a-proposta-europeia-para-regulacao-da-inteligencia-artificial

– **A Recomendação da OCDE Sobre Inteligência Artificial e sua Relevância Como Instrumento de "*Soft Law*" para Contenção dos Riscos aos Direitos Fundamentais**. Artigo apresentado como exigência parcial para obtenção do Grau de Doutoramento em Direito na disciplina "Estado, Lei e Poder Judicial – O Direito em Ação", em 2020, na Universidade Autonoma de Lisboa [*pendente de publicação*]

– **Eficiência Epistêmica Da Jurisdição Algorítmica** Artigo apresentado como exigência parcial para obtenção do Grau de Doutoramento em Direito na disciplina "Problemas Cerne", em 2019, na Universidade Autonoma de Lisboa [*pendente de publicação*]

– **Tecno-Regulação, Inteligência Artifical e Estado de Direito Democrático**. Artigo apresentado como exigência parcial para obtenção do Grau de Doutoramento em Direito na disciplina "Direito: da Norma ao procedimento e à fase aplicativa", em 2018, na Universidade Autonoma de Lisboa [*pendente de publicação*]

– **Uma Introdução ao "Direito de Informática"** [Em linha]. Brasília: IESB, 2011. [Consult. 27 mai. 2019] Disponível em https://esaoabsp.edu.br/Artigo?Art=24.

PENEL, Olivier - ***Algorithms, the Illusion of Neutrality: The Road to Trusted AI*** [Em linha]. [Consult. 20 jun. 2019] Disponível em: https://towardsdatascience.com/algorithms-the-illusion-of-neutrality-8438f9ca8471

PEREIRA, Alexandre Libório Dias – **Ius ex Machina? Da Informática Jurídica ao Computador-Juiz.** [Em linha]. [Consult. 10 jan. 2020] Disponível em: https://www.cidp.pt › revistas › rjlb

PETROPOULOS, Georgios - ***The dark side of artificial intelligence: manipulation of human behaviour*** [Em linha]. [Consult. 21 abr. 2022] Disponível em: https://www.bruegel.org/2022/02/the-dark-side-of-artificial-intelligence-manipulation-of-human-behaviour/

PETROVIĆ, Đorđe & MIJAILOVIĆ, Radomir & PEŠIĆ, Dalibor – ***Traffic Accidents with Autonomous Vehicles: Type of Collisions, Manoeuvres and Errors of Conventional Vehicles' Drivers*** [Em linha]. [Consult. 11 fev. 2021] Disponível em: https://www.sciencedirect.com/science/article/pii/S2352146520301654

PICHAI, Sundar – ***AI at Google: our principles*** [Em linha]. [Consult. 4 jul. 2019] Disponível em https://www.blog.google/technology/ai/ai-principles/

PIRES, Alex Sander Xavier - **Justiça na perspectiva Kelseniana**. Rio de Janeiro: Freitas Bastos, 2013. ISBN 978-85-7987-167-2

POOLE, David; MACKWORTH, Alan & GOEBEL, Randy (1998) - ***Computational Intelligence: A Logical Approach***. New York: Oxford University Press. ISBN 0-19-510270-3.

POP, Florina – ***Sandboxes for Responsible Artificial Intelligence***. [Em linha]. [Consult. 2 dez. 2022] Disponível em: https://www.eipa.eu/publications/briefing/sandboxes-for-responsible-artificial-intelligence/

RANCHORDAS, Sofia – ***Experimental Regulations for AI: Sandboxes for Morals and Mores*** [Em linha]. [Consult. 2 dez. 2022] Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3839744

REISMAN Dillon & SCHULTZ, Jason & CRAWFORD, Kate, WHITTAKER, Meredith - ***Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability*** [Em linha]. [Consult. 7 nov. 2022] Disponível em: *https://ainowinstitute.org › aiareport2018*

RISSE, Mathias – ***Human Rights and Artificial Intelligence*** [Em linha]. [Consult. 20 abr. 2021] Disponível em: https://www.hks.harvard.edu/publications/human-rights-and-artificial-intelligence-urgently-needed-agenda

RONSIN, Xavier – ***In-depth study on the use of AI in judicial systems, notably AI applications processing judicial decisions and data*** [Em linha]. [Consult. 8 ago. 2020] Disponível em: https://rm.coe.int › ethical-charter-en-for-publication

RUDSCHIES, Catharina & SCHNEIDER, Ingrid & SIMON, Judith – ***Value Pluralism in the AI Ethics Debate – Different Actors, Different Priorities***. [Em linha]. [Consult. 17 jun. 2022] Disponível em: https://informationethics.ca/index.php/irie/article/view/419

SANTOS, Coriolano Aurélio de Almeida Camargo & EROUD Aicha – **Inteligência artificial e direitos humanos: Uma possível dignidade da pessoa humana digital?** [Em linha]. [Consult. 15 dez. 2022] Disponível em: https://www.migalhas.com.br/coluna/direito-digital/352096/inteligencia-artificial-e-direitos-humanos

SANTY, R D et al. – ***Artificial Intelligence as Human Behavior Detection for Auto Personalization Function in Social Media Marketing***. [Em linha]. [Consult. 6 dez. 2020] Disponível em: https://ojs.unikom.ac.id › article › download

SCHERER, Matthew U. – ***Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies***. Harvard Journal of Law & Technology Volume 29, Number 2 Spring 2016. [Em linha]. [Consult. 12 nov. 2019] Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2609777

SELBST, Andrew D. **– *An Institutional View of Algorithmic Impact Assessments*** [Em linha]. [Consult. 7 nov. 2022] Disponível em: https://jolt.law.harvard.edu/assets/articlePDFs/v35/Selbst-An-Institutional-View-of-Algorithmic-Impact-Assessments.pdf

- *Regulating Inscrutable Systems*. [Em linha]. [Consult. 28 dez. 2019] Disponível em http://www.werobot2017.com/wp-content/uploads/2017/03/Selbst-and-Barocas-Regulating-Inscrutable-Systems-1.pdf

SEOANE, A. MOZO – ***La revolución tecnológica y sus retos: medios de control, falos de los sistemas y ciberdelincuencia*** [Em linha]. [Consult. 5 abr. 2021] Disponível em: https://dialnet.unirioja.es/servlet/articulo?codigo=6800583

SERVOZ, Michel – ***The future of work? The work of the future!*** [Em linha]. [Consult. 22 jun. 2022] Disponível em: https://skills4industry.eu/sites/default/files/2019-05/AI%20-%20The%20Future%20of%20Work_Work%20of%20the%20Future.pdf.

SMITH, Chris; MCGUIRE, Brian; HUANG, Ting & YANG, Gary – ***The History of Artificial Intelligence*** [Em linha]. [Consult. 8 jan. 2020] Disponível em: https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf

SOLUM, Lawrence B. – ***Legal Personhood for Artificial Intelligences***, 70 N.C. L. Rev. 1231 (1992). [Em linha]. [Consult. 9 nov. 2019] Disponível em: http://scholarship.law.unc.edu/nclr/vol70/iss4/4

STEIN, Yaakov – ***Artificial Intelligence in Social Networks Produces Polarization*** [Em linha]. [Consult. 10 ago. 2020] Disponível em: https://www.academia.edu/52211523/Artificial_Intelligence_in_Social_Networks_Produces_Polarization

SUSSKIND, Richared - ***Expert Systems in Law: a Jurisprudential Approach to Artificial Intelligence and Legal Reasoning*** [Em linha]. The Modern Law Review, 1986. [Consult. 21 dez. 2019] Disponível em: https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1468-2230.1986.tb01683.x

TEGMARK, Max - ***Life 3.0: Being Human in the Age of Artificial Intelligence***. New York: Alfred A. Knopf, 2017.

THILO, Hagendorff – ***The Ethics of AI Ethics: An Evaluation of Guidelines*** [Em linha]. [Consult. 12 ago. 2022] Disponível em: https://www.researchgate.net/publication/338983166_The_Ethics_of_AI_Ethics_An_Evaluation_of_Guidelines

TURING, Alan Mathison - ***Computing Machinery and Intelligence*** [Em linha]. Mind, a Quarterl Review of Psychology and Philosophy, Vol. LIX, nº 236, Outubro de 1950 [Consult. 23 dez. 2018] Disponível em https://academic.oup.com/mind/article-pdf/LIX/236/433/9866119/433.pdf

UBENA, John – ***The Role of Legislative Techniques in Regulation of Disruptive Technologies*** [Em linha]. [Consult. 5 jan. 2020] Disponível em http://www.calc.ngo/sites/default/files/loophole/Loophole%20-%202019-02%20%282019-06-24%29.pdf

VAN DEN BERG, Bibi – ***Robots as Tools for Techno-Regulation***. [Em linha]. [Consult. 25 nov. 2022] Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2295464

VIEIRA, Miguel Marques – **A autonomia privada na contratação electrônica sem intervenção humana.** *In* CAMPOS, Diogo Leite de (Org.) – **Estudos sobre o Direito das Pessoas**. Coimbra: Almedina, 2007, v. 1.

VOERMANS, Wim – ***Computer-assisted legislative drafting in the Netherlands: the LEDA-system*** [Em linha]. [Consult. 8 jan. 2020] Disponível em:

WADDINGTON, Matthew – ***Machine-consumable legislation: A legislative drafter's perspective – human v artificial intelligence*** [Em linha]. [Consult. 3 dez. 2021] Disponível em: https://www.researchgate.net/publication/334002003_Machine-consumable_legislation_A_legislative_drafter%27s_perspective_-_human_v_artificial_intelligence

WETTIG, Steffen and ZEHENDNER, Eberhard – ***A legal analysis of human and electronic agents*** [Em linha]. [Consult. 17 jan. 2020] Disponível em: https://www.researchgate.net/publication/220539283_A_legal_analysis_of_human_and_electronic_agents

WONG, Park-Hang – ***Cultural Differences as Excuses? Human Rights and Cultural Values in Global Ethics and Governance of AI*** [Em linha]. [Consult. 25 nov. 2022] Disponível em: https://link.springer.com/article/10.1007/s13347-020-00413-8

XU, V. X., & XIAO, B. - ***China's social credit system seeks to assign citizens scores, engineer social behaviour*** [Em linha]. [Consult. 25 dez. 2018] Disponível em http://www.abc.net.au/news/2018-03-31/chinas-social-credit-system-punishes-untrustworthy-citizens/9596204

YANKOSKI, Michael & THEISEN, William & VERDEJA, Ernesto & SCHEIRER, Walter J. – ***Artificial Intelligence for Peace: An Early Warning System for Mass Violence*** [Em linha]. [Consult. 3 mai. 2021] Disponível em: https://www.semanticscholar.org/paper/Artificial-Intelligence-for-Peace%3A-An-Early-Warning-Yankoski-Theisen/f02b9aae954aac63c3e20d7e009424cf06a097c2

YANKOSKI, Michael & WENINGER, Tim & SCHEIRER, Walter – ***An AI early warning system to monitor online disinformation, stop violence, and protect elections*** [Em linha]. [Consult. 3 mai. 2021] Disponível em: https://www.tandfonline.com/doi/full/10.1080/00963402.2020.1728976

YUDKOWSKY, Eliezer – ***Artificial Intelligence as a Positive and Negative Factor in Global Risk***. [Em linha]. [Consult. 24 jun. 2021] Disponível em: http://intelligence.org/files/AIPosNegFactor.pdf

– *Coherent Extrapolated Volition*. [Em linha]. [Consult. 14 out. 2022] Disponível em: https://intelligence.org

ZACCAGNINO, Rocco & CAPO, Carmine; GUARINO, Alfonso; e LETTIERI, Nicola & MALANDRINO, Delfina – ***Techno-regulation and intelligent safeguards Analysis of touch gestures for online child protection*** [Em linha]. [Consult. 20 dez. 2022] Disponível em: https://link.springer.com/article/10.1007/s11042-020-10446-y

ZEDNIK, Carlos - ***Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence*** [Em linha]. [Consult. 12 abr. 2020] Disponível em: https://link.springer.com/article/10.1007/s13347-019-00382-7

2. Documentais

ACCESS NOW – ***Human Rights in the Age of Artificial Intelligence***. [Em linha]. [Consult. 2 ago. 2021] Disponível em: https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf

ARIZONA STATE UNIVERSITY – ***Soft-Law Governance of Artificial Intelligence, Executive Sumary*** [Em linha]. [Consult. 12 ago. 2022] Disponível em: *https://lsi.asulaw.org/softlaw/report/executive-summary/*

BRASIL - **Lei 13.709/18 (1-05-1943)** [Em linha]. [Consult. 2 jan. 2019] Disponível em http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm

CÂMARA DOS DEPUTADOS DO BRASIL – **Projeto de Lei 21/2020** [Em linha]. [Consult. 27 set. 2022] Disponível em: *https://www.camara.leg.br/propostas-legislativas/2236340#CCTCI*

CAHAI – ***Ad hoc Committee on Artificial Intelligence - Towards Regulation Of AI*** [Em linha]. [Consult. 22 abr. 2021] Disponível em: *Systems https://www.coe.int/en/web/artificial-intelligence/cahai*

CENTER FOR SECURITY AND EMERGING TECHNOLOGY (CSET) – ***The Question of Comparative Advantage in Artificial Intelligence*** [Em linha]. [Consult. 22 abr. 2021] Disponível em: https://cset.georgetown.edu/wp-content/uploads/CSET-The-Question-of-Comparative-Advantage-in-Artificial-Intelligence-1.pdf

CHINA – ***Internet Information Service Algorithmic Recommendation Management Provisions*** [Em linha]. [Consult. 10 set. 2022] Disponível em: https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022/

COMISSÃO EUROPEIA PARA EFICIÊNCIA DA JUSTIÇA (CEPEJ) - ***European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*** [Em linha]. [Consult. 3 abr. 2019] Disponível em https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c

- ***Relatório sobre as implicações em matéria de segurança e de responsabilidade decorrentes da inteligência artificial, da Internet das coisas e da robótica*** [Consult. 22 jun. 2022] Disponível em: https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1593079180383&uri=CELEX%3A52020DC0064

COUNCIL OF EUROPE - ***Artificial Intelligence – Intelligent Politics. Challenges and opportunities for media and democracy*** [Em linha]. [Consult. 21 abr. 20212 Disponível

em: https://www.coe.int/en/web/commissioner/-/artificial-intelligence-intelligent-politics-challenges-and-opportunities-for-media-and-democracy

- **Carta Europeia de Ética no uso da inteligência artificial em sistemas judiciais e seu ambiente** [Em linha]. [Consult. 29 dez. 2018] Disponível em https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment

– ***Unboxing Artificial Intelligence: 10 steps to protect Human Rights*** [Em linha]. [Consult. 24 jun. 2021] Disponível em: https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64

GENEVA ACADEMY – ***Universality in the Human Rights Council: Challenges and Achievements*** [Em linha]. [Consult. 23 nov. 2022] Disponível em: https://www.geneva-academy.ch

EUROFOUND – ***Game-changing technologies: Transforming production and employment in Europe*** [Consult. 29 jun. 2022] Disponível em: http://eurofound.link/ef19047

EUROPEAN COMISSION – ***Artificial Intelligence for Europe*** [Em linha]. [Consult. 14 jul. 2019] Disponível em: https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe

– ***Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive)*** [Em linha]. [Consult. 29 set. 2022] Disponível em: https://eur-lex.europa.eu/procedure/EN/2022_303

– ***Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics*** [Em linha]. [Consult. 29 set. 2022] Disponível em: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0064

– ***White Paper on Artificial Intelligence - A European approach to excellence and trust*** [Em linha]. [Consult. 14 jul. 2019] Disponível em: *https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en*

EUROPEAN PARLIAMENT'S SPECIAL COMMITTEE ON ARTIFICIAL INTELLIGENCE IN A DIGITAL AGE (*AIDA*) – ***Improving working conditions using Artificial Intelligence*** [Em linha]. [Consult. 5 jun. 2020] Disponível em: https://www.europarl.europa.eu › etudes › STUD

G20 - ***Ministerial Statement on Trade and Digital Economy*** [Em linha]. [Consult. 2 jan. 2020] Disponível em: https://www.mofa.go.jp%2Ffiles%2F000486596.pdf&usg=AOvVaw328I61ihWkHj9E5fKCoqdc

GENEVA ACADEMY – ***Universality in the Human Rights Council: Challenges and Achievements*** [Em linha]. [Consult. 19 nov. 2021] Disponível em: https://www.geneva-academy.ch/research/our-clusters/past-projects/detail/14-universality-in-the-human-rights-council

HOUSE OF LORDS – ***AI in the UK: ready, willing and able?*** [Em linha]. [Consult. 3 jan. 2020] Disponível em https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

HURIGHTS OSAKA – ***Human Rights and Cultural Values: A Literature Review*** [Em linha]. [Consult. 30 nov. 2022] Disponível em: https://www.hurights.or.jp/archives/database/hr-cultural-values.html

IBA GLOBAL EMPLOYMENT INSTITUTE - ***Artificial Intelligence and Robotics and Their Impact on the Workplace***. [Consult. 24 jun. 2022] Disponível em: https://onwork.edu.au/bibitem/2017-Wisskirchen,G-Biacabe,B+T-etal-Artificial+intelligence+and+robotics+and+their+impact+on+the+workplace/

IBM – ***IBM's Principles for Trust and Transparency*** [Em linha]. [Consult. 15 jan. 2020] Disponível em https://www.ibm.com/blogs/policy/trust-principles/

INSTITUTE OF ELECTRICAL AND ELETRONICAL ENGINEERS (*IEEE*) – ***Ethically Aligned Design.*** [Em linha]. [Consult. 3 jan. 2020] Disponível em https://standards.ieee.org/industry-connections/ec/autonomous-systems.html

INTERNATIONAL COMMITTEE OF THE RED CROSS (IRCR) – ***Autonomy, artificial intelligence and robotics: Technical aspects of human control***. [Em linha]. [Consult. 19 fev. 2021] Disponível em: https://www.icrc.org/en/document/autonomy-artificial-intelligence-and-robotics-technical-aspects-human-control

– ***ICRC Position on Autonomous Weapon Systems*** [Em linha]. [Consult. 26 jan. 2022] Disponível em: https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems

INTERNATIONAL LABOUR OFFICE – ***Report on collective dismissals - A comparative and contextual analysis of the law on collective redundancies in 13 European countries*** [Consult. 29 jun. 2022] Disponível em: https://www.ilo.org › publication › wcms_541637

INTERNATIONAL LABOUR ORGANIZATION – EMPLOYMENT POLICY DEPARTMENT – ***"Negotiating the algorithm": Automation, artificial intelligence and labour protection***. [Em linha]. [Consult. 26 jun. 2022] Disponível em: https://www.ilo.org/employment/Whatwedo/Publications/working-papers/WCMS_634157/lang--en/index.htm

Report on collective dismissals – A comparative and contextual analysis of the law on collective redundancies in 13 European countries [Em linha]. [Consult. 9 mar. 2021] Disponível em: https://labordoc.ilo.org/discovery/fulldisplay/alma994932093002676/41ILO_INST:41ILO_V1

MICROSOFT – ***Microsoft AI principles*** [Em linha]. [Consult. 18 jan. 2020] Disponível em: https://www.microsoft.com/en-us/ai/our-approach-to-ai

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (*OECD*)

– ***Artificial Intelligence in Society*** [Em linha]. [Consult. 8 jul. 2019] Disponível em: https://doi.org/10.1787/eedfee77-en

– ***Recommendation of the Council on Artificial*** (OECD/LEGAL/0449) [Em linha]. [Consult. 8 jul. 2019] Disponível em: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

– ***The impact of Artificial Intelligence on the labour market: What do we know so far?*** [Em linha]. [Consult. 23 out. 2021] Disponível em: https://www.oecd-ilibrary.org/social-issues-migration-health/the-impact-of-artificial-intelligence-on-the-labour-market_7c895724-en

ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD) – ***The impact of Artificial Intelligence on the labour market: What do we know so far?*** [Em linha]. [Consult. 12 mai. 2021] Disponível em: https://www.oecd-ilibrary.org/social-issues-migration-health/the-impact-of-artificial-intelligence-on-the-labour-market_7c895724-en

OXFORD INSIGHTS – ***Government AI Readiness Index 2021*** [Em linha]. [Consult. 21 abr. 20212] Disponível em: https://www.oxfordinsights.com/government-ai-readiness-index2021

NATURE INDEX – ***The race to the top among the world's leaders in artificial intelligence***. [Em linha]. [Consult. 21 abr. 20212] Disponível em: https://www.nature.com/articles/d41586-020-03409-8

PARLAMENTO EUROPEU – **Recomendações à Comissão sobre disposições de Direito Civil sobre Robótica** [Em linha]. [Consult. 8 jul. 2019] Disponível em http://www.europarl.europa.eu/doceo/document/A-8-2017-0005_PT.html#title2

– ***Improving working conditions using Artificial Intelligence*** [Consult. 06 jun. 2022] Disponível em: https://www.europarl.europa.eu/committees/en/aida/home/highlights

STATE OF CALIFORNIA - ***Assembly Concurrent Resolution No. 215*** (Sept. 7, 2018) [Em linha]. [Consult. 8 jan. 2020] Disponível em https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180ACR215

THE LAW LIBRARY OF CONGRESS – ***Regulation of Artificial Intelligence: The Americas and the Caribbean*** [Em linha]. [Consult. 8 out. 2022] Disponível em: *https://tile.loc.gov › storage-services › service › llglrd*

UNITED NATIONS (*UN*) – ***Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*** [Em linha]. [Consult. 23 jun. 2019] Disponível em https://undocs.org/A/73/150

– ***The impact of the technological revolution on labor markets and income distribution*** [Em linha]. [Consult. 12 out. 2021] Disponível em: https://www.un.org › 2017_Aug_Frontier-Issues-1

– ***United Nations Activities on Artificial Intelligence (AI) 2021*** [Em linha]. [Consult. 5 dez. 2022] Disponível em: https://aiforgood.itu.int/about-ai-for-good/un-ai-actions/

UNITED STATES CONGRESS - ***National Artificial Intelligence Initiative Act of 2020*** [Em linha]. [Consult. 29 set. 2022] Disponível em: https://www.congress.gov/bill/116th-congress/house-bill/6216/text

- ***Algorithmic Accountability Act of 2022*** [Em linha]. [Consult. 29 set. 2022] Disponível em: https://www.congress.gov/bill/117th-congress/house-bill/6580/text

WORLD INTELLECTUAL PROPERTY ORGANIZATION - ***WIPO Technology Trends 2019 Executive summary Artificial Intelligence*** [Em linha]. [Consult. 1 set. 2022] Disponível em: *https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055_exec_summary.pdf*